

SAULO HENRIQUE WEBER

**Desenvolvimento de nova função densidade de probabilidade
para avaliação de regeneração natural**

Dissertação apresentada como requisito parcial para a obtenção do título de Mestre em Engenharia Florestal no Programa de Pós-Graduação em Engenharia Florestal da Universidade Federal do Paraná.

Orientador: Prof. Dr. Julio Eduardo Arce
Co-orientador: Prof. Dr. Sylvio Péllico Netto

CURITIBA
2006

SUMÁRIO

1 - INTRODUÇÃO	1
JUSTIFICATIVAS	4
OBJETIVOS	4
Objetivo Geral	4
Objetivos Específicos	5
2 - REVISÃO DE LITERATURA	6
REGENERAÇÃO NATURAL	6
SASSAFRÁS	8
TÓPICOS DE MATEMÁTICA	8
Limites	10
Derivadas	11
Integrais	11
Integrais impróprias	12
MODELAGEM	12
TÓPICOS DE ESTATÍSTICA	14
DISTRIBUIÇÕES	16
PRINCIPAIS DISTRIBUIÇÕES DISCRETAS	18
PRINCIPAIS DISTRIBUIÇÕES CONTÍNUAS	19
DISTRIBUIÇÃO EXPONENCIAL	20
DISTRIBUIÇÃO GAMMA	21
DISTRIBUIÇÃO BETA	22
DISTRIBUIÇÃO WEIBULL	25
DISTRIBUIÇÃO NORMAL OU DE GAUSS	27
3 - METODOLOGIA	30
LOCALIZAÇÃO DA ÁREA	30
COLETA DE DADOS	31
AJUSTE DOS MODELOS	33
MÉDIA E VARIÂNCIA DE UMA FUNÇÃO DENSIDADE DE PROBABILIDADE	34
EXPECTATIVA MATEMÁTICA	34
AVALIAÇÃO DO AJUSTE	34
□ COEFICIENTE DE DETERMINAÇÃO MÚLTIPLO	35
□ COEFICIENTE DE DETERMINAÇÃO MÚLTIPLO AJUSTADO	35

□ ERRO PADRÃO DA ESTIMATIVA RECALCULADO	35
□ KOLMOGOROV E SMIRNOV	36
□ QUI-QUADRADO (χ^2).....	36
□ ANÁLISE DE RESÍDUOS	36
□ DESVIO RELATIVO	37
4 – RESULTADOS E DISCUSSÃO - DESENVOLVIMENTO DA FUNÇÃO DENSIDADE DE PROBABILIDADE PROPOSTA	38
CONDIÇÕES PARA QUE UMA FUNÇÃO SEJA CONSIDERADA UMA FDP:	42
MÉDIA ARITMÉTICA	43
VARIÂNCIA.....	43
MODA	44
PONTOS DE INFLEXÃO	44
CONDIÇÕES DE EXISTÊNCIA DA FUNÇÃO DE DENSIDADE DE PROBABILIDADE.....	46
TENDÊNCIA DO MODELO DESENVOLVIDO A NORMALIDADE.....	46
FORMAS QUE A FUNÇÃO PODE ASSUMIR.....	47
5 - SÍNTESE DO MODELO PROPOSTO	54
6 - APLICAÇÃO/VERIFICAÇÃO	55
NORMAL.....	56
WEIBULL	56
GAMMA	56
BETA.....	56
EXPONENCIAL	56
MODELO PROPOSTO.....	56
7 - CONCLUSÕES.....	60
8 - REFERÊNCIAS	62
ANEXO 1 – Linhas de comando para o desenvolvimento da Função densidade de probabilidade.	66
FUNÇÃO DENSIDADE DE PROBABILIDADE	66
Função	66
Propriedades da função.....	66
Função probabilística.....	66
ANEXO 2 – Linhas de comando para a obtenção das expressões que estimam a média, variância e desvio padrão.....	69
Expectativa Matemática	69

Média	69
Variância.....	70
Desvio padrão	71
ANEXO 3 – Linhas de comando para a obtenção das expressões que estimam a moda e os pontos de inflexão.	72
Função	72
Moda.....	72
Pontos de inflexão	72
ANEXO 4 – Procedimentos para a obtenção dos coeficientes, utilizando-se o software Table curve 2D.....	74

LISTA DE FIGURAS

FIGURA 1 – REPRESENTAÇÃO DA ÁREA DEFINIDA ENTRE A CURVA E O EIXO DAS ABSCISSAS	12
FIGURA 02 – REPRESENTAÇÃO DA ÁREA SOB A CURVA	17
FIGURA 03 – DISTRIBUIÇÃO EXPONENCIAL, QUANDO $\beta = 1$	20
FIGURA 04 - REPRESENTAÇÃO GRÁFICA DA DISTRIBUIÇÃO GAMMA, PARA $a=1$ (SÓLIDO), $a=1,5$ (PONTOS), $a=2$ (TRAÇOS) E $a=3$ (PONTOS E TRAÇOS) E $b=0,5$	22
FIGURA 05 – REPRESENTAÇÃO GRÁFICA DA DISTRIBUIÇÃO BETA, PARA $a=1$ E $b = 2$ (SÓLIDO), $a=2$ E $b = 1$ (PONTOS) E $a=1$ E $b = 1$ (TRAÇOS)	23
FIGURA 06 – REPRESENTAÇÃO GRÁFICA DA DISTRIBUIÇÃO BETA, PARA $a=0,3$ E $b = 0,8$ (SÓLIDO) E $a=0,3$ E $b = 0,3$ (PONTOS)	24
FIGURA 07 – REPRESENTAÇÃO GRÁFICA DA DISTRIBUIÇÃO BETA, PARA $a=2$ E $b = 3$ (SÓLIDO), $a=5$ E $b = 3$ (PONTOS) E $a=2$ E $b = 2$ (TRAÇOS)	24
FIGURA 08 – REPRESENTAÇÃO GRÁFICA DA DISTRIBUIÇÃO BETA, PARA $a = 0,5$ E $b = 1,5$ (SÓLIDO) E $a = 3$ E $b = 0,5$ (PONTOS)	25
FIGURA 09 - REPRESENTAÇÃO GRÁFICA DA DISTRIBUIÇÃO WEIBULL, COM $a = 1$, $b = 10$ e $c = 0,7$ (SÓLIDO), COM $a = 1$, $b = 10$ e $c = 1$ (PONTOS), COM $a = 1$, $b = 10$ e $c = 2$ (TRAÇOS), COM $a = 1$, $b = 10$ e $c = 3,6$ (PONTOS E TRAÇOS)	26
FIGURA 10 – DISTRIBUIÇÃO NORMAL, $\sigma = 1$ E $\mu = 0$	27
FIGURA 11 – LOCALIZAÇÃO DA FAZENDA EXPERIMENTAL GRALHA AZUL	30
FIGURA 12 – MAPA DA ÁREA DA UNIDADE EXPERIMENTAL	31
FIGURA 13 - REPRESENTAÇÃO DAS PARCELAS PERMANENTES.	32
FIGURA 14 – REPRESENTAÇÃO GRÁFICA DA EQUAÇÃO 27 AJUSTADO AOS DADOS DE ALTURA DE REGENERAÇÃO NATURAL DE SASSAFRÁS	38
FIGURA 15 – REPRESENTAÇÃO GRÁFICA DA EQUAÇÃO 28 AJUSTADO AOS DADOS DE ALTURA DE REGENERAÇÃO NATURAL DE SASSAFRÁS	39
FIGURA 16 - REPRESENTAÇÃO GRÁFICA DA EQUAÇÃO 29 AJUSTADO AOS DADOS DE ALTURA DE REGENERAÇÃO NATURAL DE SASSAFRÁS	40

FIGURA 17 - REPRESENTAÇÃO GRÁFICA DA EQUAÇÃO 30 AJUSTADO AOS DADOS DE ALTURA DE REGENERAÇÃO NATURAL DE SASSAFRÁS	40
FIGURA 18 - REPRESENTAÇÃO GRÁFICA DA EQUAÇÃO 31 AJUSTADO AOS DADOS DE ALTURA DE REGENERAÇÃO NATURAL DE SASSAFRÁS	41
FIGURA 19 – REPRESENTAÇÃO GRÁFICA DO MODELO PROPOSTO, VARIANDO-SE OS COEFICIENTES, DE ACORDO COM AS CONCLUSÕES DO ESTUDO SOBRE OS PONTOS DE INFLEXÃO	45
FIGURA 20 – REPRESENTAÇÃO TRIDIMENSIONAL DAS FORMAS GRÁFICAS ASSUMIDAS PELO MODELO PROPOSTO	48
FIGURA 21 – REPRESENTAÇÃO GRÁFICA DO MODELO PROPOSTO, EM QUE OS COEFICIENTES SÃO ARBITRARIAMENTE CONSIDERADOS IGUAIS.	48
FIGURA 22 – REPRESENTAÇÃO GRÁFICA DO MODELO PROPOSTO, COM a , c E d FIXOS E VARIANDO b	49
FIGURA 23 – REPRESENTAÇÃO GRÁFICA DO MODELO PROPOSTO, COM a , b E d FIXOS E VARIANDO c	50
FIGURA 24 – REPRESENTAÇÃO GRÁFICA DO MODELO PROPOSTO, COM b , c E d FIXOS E VARIANDO a	51
FIGURA 25 – REPRESENTAÇÃO GRÁFICA DO MODELO PROPOSTO, COM a , b E c FIXOS E VARIANDO d	52
FIGURA 26 – REPRESENTAÇÃO GRÁFICA DAS FUNÇÕES CLÁSSICAS E A DESENVOLVIDA NESTE TRABALHO AJUSTADAS AOS DADOS DE REGENERAÇÃO NATURAL DE SASSAFRÁS E AS RESPECTIVAS FREQUÊNCIAS OBSERVADAS	57
FIGURA 27 – RESÍDUOS EM PORCENTAGEM DAS FUNÇÕES CLÁSSICAS E A DESENVOLVIDA NESTE TRABALHO AJUSTADAS AOS DADOS DE REGENERAÇÃO NATURAL DE SASSAFRÁS	58

LISTA DE EQUAÇÕES

EQUAÇÃO 01.....	20
EQUAÇÃO 02.....	21
EQUAÇÃO 03.....	21
EQUAÇÃO 04.....	21
EQUAÇÃO 05.....	22
EQUAÇÃO 06.....	22
EQUAÇÃO 07.....	22
EQUAÇÃO 08.....	25
EQUAÇÃO 09.....	25
EQUAÇÃO 10.....	25
EQUAÇÃO 11.....	26
EQUAÇÃO 12.....	26
EQUAÇÃO 13.....	26
EQUAÇÃO 14.....	27
EQUAÇÃO 15.....	27
EQUAÇÃO 16.....	27
EQUAÇÃO 17.....	34
EQUAÇÃO 18.....	34
EQUAÇÃO 19.....	35
EQUAÇÃO 20.....	35
EQUAÇÃO 21.....	35
EQUAÇÃO 22.....	35
EQUAÇÃO 23.....	36
EQUAÇÃO 24.....	36
EQUAÇÃO 25.....	36
EQUAÇÃO 26.....	37
EQUAÇÃO 27.....	38
EQUAÇÃO 28.....	38
EQUAÇÃO 29.....	39
EQUAÇÃO 30.....	40
EQUAÇÃO 31.....	41
EQUAÇÃO 32.....	41

EQUAÇÃO 33.....	42
EQUAÇÃO 34.....	43
EQUAÇÃO 35.....	43
EQUAÇÃO 36.....	44
EQUAÇÃO 37.....	44
EQUAÇÃO 38.....	56
EQUAÇÃO 39.....	56
EQUAÇÃO 40.....	56
EQUAÇÃO 41.....	56
EQUAÇÃO 42.....	56

LISTA DE TABELAS

TABELA 01 – DISTRIBUIÇÃO DE FREQUÊNCIAS DA ALTURA DA REGENERAÇÃO NATURAL DE SASSAFRÁS	55
TABELA 02 – COEFICIENTES AJUSTADOS.....	55
TABELA 03 – MÉDIA, DESVIO PADRÃO E COEFICIENTE DE VARIAÇÃO.....	59
TABELA 04 – TESTES DE ADERÊNCIA E COMPARAÇÃO ENTRE FUNÇÕES.....	59

RESUMO

Ao pesquisar diversas espécies arbóreas contidas nos fragmentos de floresta nativa remanescentes na Fazenda Experimental Gralha Azul da Pontifícia Universidade Católica do Paraná - PUCPR, localizada no município de Fazenda Rio Grande, PR, verificou-se que as distribuições probabilísticas aplicadas atualmente não são suficientemente flexíveis para representar a frequência de sobreviventes da regeneração natural. A partir da análise dos dados coletados, verificou-se que a quantidade de sementes que germinam no interior da floresta é grande, porém, a porcentagem de plantas que se tornam adultas (DAP maior que 10 cm), em relação à quantidade de sementes que germinam, é relativamente baixa. Essa situação não é bem representada por nenhuma das distribuições clássicas, como a Normal, a Weibull, a Gamma, a Beta e a Exponencial, que são comumente utilizadas para esses casos. Fundamenta-se este trabalho na identificação de um modelo matemático que atenda às seguintes condições: ser suficientemente flexível para representar a distribuição de frequências das alturas de regeneração natural de espécies nativas; e que possa ser convertida em uma função densidade de probabilidade, atendendo às características e propriedades atinentes a ela. Este trabalho teve como objetivo principal testar modelos matemáticos, que foram ajustados por meio de regressão não linear, segundo o procedimento de Marquadt, pelo método de mínimos quadrados e com múltiplas iterações. A característica principal dessa função é a flexibilidade para representar a grande quantidade de indivíduos contidos na primeira classe, seguido de um decréscimo abrupto nas classes subseqüentes. Esse modelo resultou em um melhor ajuste à distribuição de alturas de regeneração natural, em comparação com os modelos clássicos. Foram desenvolvidas as fórmulas para estimar os parâmetros média, variância e moda, bem como para a obtenção dos pontos de inflexão do modelo. As formas que a função pode assumir foram simuladas de acordo com a variação dos coeficientes. Ao comparar as principais distribuições utilizadas no meio florestal e a desenvolvida neste trabalho, verificou-se que essa representou melhor os dados, resultou em bons valores nos testes de aderência, podendo inclusive ser utilizada como função preditiva. A primeira parte do presente trabalho consistiu em revisar a literatura pertinente para aprofundar conhecimentos sobre as distribuições clássicas utilizadas para avaliar mortalidade de regeneração natural. A seguir, manipulando dados de regeneração natural de Sassafrás (*Ocotea odorifera* (Vell.) Rowher), encontrou-se que os modelos probabilísticos clássicos (Normal, Weibull, Gamma, Beta e Exponencial) não resultaram em um bom ajuste a esses dados. Em um terceiro momento, foi tentado conseguir novas alternativas para resolver este problema, buscando várias possibilidades aplicativas de outras funções matemáticas que pudessem representar bem a dispersão de dados. Após definido o novo modelo procedeu-se às operações para adequá-lo aos requisitos de uma Função Densidade de Probabilidade (FDP). O quinto passo foi derivar as fórmulas para calcular a média e a variância, usando-se o método de expectativa matemática. Na seqüência, foram derivadas as fórmulas para a obtenção da moda e dos pontos de inflexão, bem como foram simuladas as formas das curvas, usando-se variações dos coeficientes da função. Finalmente, os resultados das funções clássicas ajustadas e os do modelo proposto na presente pesquisa foram comparados, usando-se os testes de aderência. O modelo proposto no presente trabalho mostrou-se ser suficientemente flexível para o ajuste aos dados de altura da regeneração natural do Sassafrás. A média aritmética, o desvio padrão e a moda, calculados usando as fórmulas derivadas do modelo proposto, resultaram em 2,29 m, 2,71 m e 0,8 m, respectivamente, os quais tiveram muito boas aproximações com os valores estimados da amostra (2,29 m, 2,66 m e 1,04 m).

Palavras-chave: Distribuições Probabilísticas Assimétricas; Regressão Não Linear; Florestas Naturais; Mortalidade; Regeneração Natural.

ABSTRACT

Development of a new probability density function to evaluate natural regeneration

Doing a research in remaining fragments of the Araucaria Forest in the Gralha Azul Experimental Farm of the Pontifical Catholic University of Parana – PUCPR, located at the municipality of Fazenda Rio Grande, PR, it was observed various forest species, where it was verified that the probability distributions currently used does not satisfy the flexibility required to well represent the frequency of natural regeneration survivors. From an analysis of data collected, it was verified that the amount of germinated seeds was high, however, the percentage of these plants that became mature (DBH bigger than 10 cm), in relation to the amount of germinated seeds, was relatively low. This situation is not represented satisfactorily by any of the classic probabilistic distributions as Normal, Weibull, Gamma, Beta and Exponential, which are commonly used for these cases. This work was based on the necessity of identifying a mathematical model that includes the following conditions: presents enough flexibility to represent the frequency distribution of natural regeneration; and that can be converted into a probability density function, considering its characteristics and properties. This work had as main objective to test mathematical models that was adjusted by means of non-linear regression, Marquadt procedure, by the least squares method, with multiple iterations. The main characteristic of this function was to attain enough flexibility to represent the great amount of individuals in the first class, followed by an abrupt decrease in the following classes. This model resulted in a better adjustment for the variable height of natural regeneration data, in comparison to the application of the classical models. The frequency distribution was better represented by the proposed model in this work and resulted in good approximation of the estimated mean and variance values. The expressions were developed to calculate these parameters, as well as those to calculate the mode and the inflection points. The shapes that the function can assume were simulated in accordance with the parameters variations. When comparing the main distributions used in the forest science and the one developed in this work, it was verified that it showed a better adjustment to the height data, also viable to be used as predictive function, and presented non-significance when the application of the statistical adjustment tests. The first part of the present work consisted of literature revision to be acquainted with the application of the classical distributions to evaluate mortality of natural regeneration. After manipulating natural regeneration data of Sassafras (*Ocotea odorifera* (Vell.) Rowher), it was verified that the classical probabilistic functions (Normal, Weibull, Gamma, Beta and Exponential) did not well adjusted to mortality and survival distribution data, because the shape of the observed data required a better flexibility of the mathematical model. At the third moment, it was tried to get new alternatives to solve this problem, by searching various possible applications of other mathematical functions that could well represent the data dispersion. After defined the new proposed model, it was proceeded operations to adequate it to the requirements of a Probability Density Function (PDF). The fifth step was to derive the formulas to calculate the mean and the variance, using the method of mathematical expectation. After, the mode and the inflection points formulas were derived, as well as the shape curves was simulated, using variation of the parameters values. Finally, the results of the adjusted classical functions and the proposed in the present research were compared by means of goodness of fit tests. The model proposed in this research work showed to be enough flexible for the adjustment of Sassafras natural regeneration height data. The mean, the standard deviation and the mode

calculated by the formulas derived from the proposed distribution model resulted in 2.29 m, 2.71 m and 0,8m, respectively, which had very good approximation with the estimated sample values (2.29 m, 2.66 m and 1,04m).

Keywords: Asymmetrical Probabilistic distributions; Non-linear Regression; Nature Forests; Mortality; Natural Regeneration.

1 - INTRODUÇÃO

A amostragem é uma ferramenta de suma importância para se conhecer populações e é por meio dessa que são calculadas as estimativas dos seus parâmetros. O desenvolvimento dessa área do conhecimento está intimamente relacionado com a evolução da estatística, auxiliados pela matemática.

Será abordada, a seguir, uma breve discussão sobre o desenvolvimento dos censos e inventários ocorridos ao longo da história.

O Egito antigo, cuja sobrevivência da população dependia principalmente da agricultura, desenvolveu-se às margens do Rio Nilo e possuía um programa de distribuição de alimentos por meio da cobrança de impostos. Sabe-se que esses impostos eram cobrados proporcionalmente à área e que a cada ano ocorria a cheia do Rio Nilo, desmarcando as divisas que localizavam as propriedades. Portanto, para efetuar a cobrança dessas taxas, era necessário fazer um censo para conhecer a área cultivável total e quanto dessas terras cada agricultor possuía (STRUICK, 1987).

Na Grécia antiga surgiram os primeiros estudos demográficos dos quais se tem conhecimento. Nessas pesquisas, a população toda era entrevistada, uma vez que não existia a idéia de amostragem. Com o passar dos anos e o aumento da população, tornou-se extremamente trabalhoso fazer esse levantamento, necessitando-se, assim, de um sistema mais rápido e que demandasse uma menor quantidade de mão-de-obra (IFRAH, 1947).

Um censo envolve um exame de todos os elementos de um dado grupo, ao passo que a amostragem envolve o estudo de apenas uma parte dos elementos. A finalidade da amostragem é fazer generalizações sobre todo um grupo sem precisar examinar cada um de seus elementos.

Segundo BOYER (1996), a palavra estatística surgiu da expressão em Latim *statisticum collegium* (palestra sobre os assuntos do Estado), originando o termo em língua italiana *statista*, que significa "homem de estado", ou político. Em alemão, *Statistik* designa a análise de dados sobre o Estado. O sentido adotado atualmente surgiu no início do século 19.

Salienta-se que o desenvolvimento de certas áreas da experimentação biológica só foi possível devido ao uso dos métodos estatísticos (MOOD; GRAYBILL, 1978). Foram então esses mesmos biólogos os que desenvolveram a parte essencial desta teoria.

A definição de estatística mais habitual atualmente é que essa é uma área do conhecimento que utiliza teorias probabilísticas para explicação de eventos, estudos e experimentos. Essa teoria tem por objetivo obter, organizar e analisar dados, determinar suas

correlações, tirar delas suas consequências para descrição, explicar o que passou e fazer previsões.

A estatística pode ser descritiva, em que o objetivo é, em geral, a simplificação de informações, de probabilidade, situações em que envolvam acaso, e de inferência, que é a análise e interpretação de dados amostrais (JOHNSON; LEONE, 1964).

De acordo com EUES (1995), o passo decisivo para fundamentação teórica da inferência estatística, associa-se ao desenvolvimento do cálculo das probabilidades, que teve suas origens nos jogos de azar, no século XVII. Há autores que sustentam que o cálculo das probabilidades teve a sua origem na Itália com Paccioli (1445-1514), Cardano (1501-1576), Tartaglia (1499-1557), Galileo (1564-1642) e outros. O livro "De Ratiociniis in Ludo Aleae", escrito por Huyghens (1629-1645), é considerado o primeiro livro sobre cálculo das probabilidades e tem a particularidade notável de introduzir o conceito de esperança matemática. Nessa publicação ele cita que as probabilidades são utilizadas para exprimir a chance de ocorrência de determinado evento.

A amostragem é o ato de obter uma amostra de uma população da qual se deseja fazer inferência. O levantamento por amostragem, quando comparado com o censo, apresenta custos menores, resultados em menor tempo e objetivos mais amplos. Uma amostragem probabilística ocorre quando cada elemento da população, da qual é escolhida a amostra, tem uma probabilidade conhecida e diferente de zero de pertencer a ela.

Com objetivo de ampliar os estudos e aprofundar os resultados, foram desenvolvidas distribuições de frequência, que são grupamentos de dados em classes. Uma distribuição de frequências relativas, para os resultados de um espaço amostral, constitui uma distribuição de probabilidade. Essas distribuições podem ser contínuas ou descontínuas, de acordo com os tipos de variáveis em questão, discretas ou contínuas. Enquanto uma variável aleatória discreta pode ser escolhida de possíveis valores finitos ou infinitos numeráveis, as variáveis aleatórias contínuas apresentam um conjunto de possibilidades incontáveis ou não numeráveis.

Com a criação da teoria de amostragem houve a necessidade do desenvolvimento de Funções Probabilísticas para a avaliação de distribuições contínuas. Assim, surgiram as Funções Densidade de Probabilidade. Segundo ROSS (1988), a primeira e mais utilizada função desenvolvida foi a Curva Normal, criada pelo matemático francês Abraham De Moivre, em 1733, com a finalidade de aproximar probabilidades referentes a lançamentos de moedas. Entretanto, apenas em 1809 o conceituado matemático Johann Carl Friedrich Gauss a usou como uma parte integral da sua aproximação para prever o local de entidades

astronômicas, publicada na terceira seção do livro *Theoria Motus*. Nesse modelo a área sob a curva entre a média e um ponto arbitrário é função do número de desvios padrões entre a média e aquele ponto. Essa distribuição desempenha um papel essencial na história do estudo das populações e é amplamente aplicado na área biológica, resultando em uma excelente aderência. Foi com base nesta função que Gauss desenvolveu a teoria de regressão linear, fundamental para a evolução do ajuste de modelos preditivos.

Novas pesquisas, nas mais diversas áreas, foram realizadas e foram descobertas distribuições de dados que não eram satisfatoriamente representados pela Curva Normal, o que fomentou a busca por distribuições assimétricas em relação à média aritmética.

Dos modelos mais utilizados, pode-se citar a Função Weibull, desenvolvida em Stockholm por Ernst Hjalmar Waloddi Weibull em 1939 com a finalidade de realizar um planejamento estatístico sobre fadiga de materiais. A distribuição Beta é utilizada principalmente para análises Bayesianas, PERT e CPM, por ser binomial. A função Gamma é chamada de distribuição χ_n^2 (lê-se qui-quadrado) com n graus de liberdade quando $\lambda = \frac{1}{2}$ e

$t = \frac{n}{2}$, n sendo um inteiro positivo. A Exponencial pode ser considerada como caso particular da Gamma e tem sido utilizada para prever o período de tempo necessário até a ocorrência de um evento. No entanto, mesmo essas distribuições vêm tendo desempenho abaixo das expectativas quando a variabilidade dos dados é alta.

BARROS (1980), no interior da Amazônia, pesquisando a distribuição diamétrica de espécies arbóreas nativas, encontrou-se um conjunto de dados de forma platicúrtica, diferente de tudo que se conhecia nesse âmbito. Nesse estudo foram ajustados dois tipos de funções Exponencial, bem como os modelos de Mervart, Pierlot, Goff e West, Weibull e Beta.

As florestas naturais do Brasil possuem características únicas em exuberância e heterogeneidade. Há um esforço por parte dos pesquisadores para entender o comportamento do crescimento de espécies nativas, tanto individualmente, quanto coletivamente. Esses profissionais vêm buscando na estatística novas funções que sejam suficientemente flexíveis para aplicá-las em dados de regeneração natural dessas espécies.

Outro ponto importante a ser considerado é o tamanho das populações florestais. Essas geralmente se compõem de extensas áreas e a única alternativa para abordá-las é a amostragem, uma vez que nessas circunstâncias torna-se inviável o censo, em função do tempo e, principalmente, do custo requerido para essa operação.

Muitos estudos têm sido realizados no Brasil a fim de entender a dinâmica de crescimento de espécies nativas. Nessas pesquisas tem se utilizados diversas funções densidade de probabilidade, além das citadas anteriormente, como por exemplo, a Log-Normal, que é uma variação da normal.

JUSTIFICATIVAS

Uma função matemática que possa atender às avaliações de mortalidade e sobrevivência de regeneração natural terá importância e aplicação para a elaboração dos planos de manejo de florestas naturais, bem como conhecer com maior profundidade a tendência de mortalidade ocorrente nesses ambientes.

Após aplicar os modelos clássicos à distribuição de alturas do Sassafrás (*Ocotea odorifera*) verificou-se que esses não resultaram em bom ajuste, principalmente nas primeiras classes, onde estão concentradas até 90% dos indivíduos.

Atualmente, o modelo mais utilizado para representação da regeneração natural de espécies nativas é o exponencial, conhecido na literatura florestal como “J invertido”. Esse modelo apresenta um bom ajuste, mas considera, em uma distribuição de alturas, que há β indivíduos com “altura zero”, o que não é uma realidade.

Fundamenta-se esta pesquisa na necessidade de desenvolver uma função densidade de probabilidade capaz de representar, de forma satisfatória, a distribuição de alturas de espécies arbóreas nativas. Esse modelo deve ser suficientemente flexível para representar a grande variação ocorrente entre as classes, ou seja, a grande quantidade de indivíduos contida nas primeiras classes e o pequeno número de regenerações que formam as classes subseqüentes, bem como passar pela origem, ou seja, quando $x = 0$, $f(x) = 0$. Admitindo-se essa última premissa supõe-se teoricamente que não existem indivíduos com altura 0 (zero).

OBJETIVOS

Objetivo Geral

O objetivo principal desta pesquisa foi ajustar modelos matemáticos que possam ser convertidos em funções densidade de probabilidade e que sejam capazes de permitir a avaliação probabilística da regeneração natural de espécies florestais nativas.

Objetivos Específicos

- Testar diversas funções, a fim de escolher o modelo que melhor representa a distribuição dos dados de regeneração natural;
- Comparar a eficiência desta e de outras funções densidade de probabilidade clássicas, como Normal, Weibull, Gamma, Beta e Exponencial, que atualmente são as mais aplicadas para essa caracterização;
- Testar o modelo matemático selecionado em distribuição de alturas de uma espécie nativa para comprovar sua eficiência em termos de flexibilidade e aderência;
- Desenvolver as fórmulas para calcular os parâmetros média, variância e moda, bem como para os pontos de inflexão; e
- Simular as formas que a função poderá assumir, de acordo com a variação dos coeficientes.

2 - REVISÃO DE LITERATURA

REGENERAÇÃO NATURAL

A regeneração natural é um processo importante para a sobrevivência das florestas. Para SEITZ (1994) esse é um processo característico de cada espécie, em perfeita sintonia com as condições ambientais e do meio, permitindo, assim, a perpetuação do patrimônio genético durante o tempo.

Para CRESTANA (2000), a regeneração natural substitui as árvores mortas na floresta. CORVELLO (1983) diz que a regeneração natural constitui um alicerce para a sobrevivência e o desenvolvimento do ecossistema florestal e representa um desafio para os técnicos da área florestal devido, em parte, aos escassos conhecimentos sobre auto-ecologia das espécies nativas.

Considerando a dinâmica de crescimento e desenvolvimento das florestas, o estudo da composição e da estrutura da regeneração natural é obrigatório para a elaboração e aplicação correta dos planos de manejo silvicultural, permitindo, assim, o aproveitamento racional e permanente de áreas florestais (CARVALHO, 1982).

Segundo MARIANO et al. (1998), a regeneração natural de espécies clímax resulta do banco de plântulas ou da queda de sementes em locais propícios ao seu desenvolvimento. SEITZ (1994) afirmou que todas as espécies possuem condições de se regenerar naturalmente, entretanto o sucesso dessa dinâmica está diretamente relacionado a vários fatores, tais como: produção, dispersão e sanidade de sementes / propágulos, presença de predadores, umidade do substrato, temperatura, inibidores bioquímicos, energia (luz), nutrientes e micorrizas. O sucesso da regeneração natural depende da germinação e competição, pois no fim do curto processo da germinação a plântula exibe uma superfície fotossintética suficiente para o seu próprio abastecimento em alimento. A competição é um processo que define a intensidade de regeneração por espécie, em que as plantas sobreviventes serão aquelas que estiverem genética e fisiologicamente adaptadas a utilizar ao máximo os fatores de produção (INOUE, 1979).

Segundo SANQUETA; NINOMIYA; OGINO (1992), os tipos de estratégia de regeneração natural de florestas nativas podem ser basicamente os seguintes:

- Espécies formadoras de banco de sementes: pioneiras depositoras de sementes que permanecem dormentes no solo à espera por distúrbios na floresta que favoreçam a sua germinação;

- Espécies formadoras de banco de mudas: espécies clímax produtoras de sementes que germinam e sobrevivem sob a sombra;
- Espécies dispersoras: espécies produtoras de sementes que poderão germinar sob a sombra, mas que não sobrevivem de maneira significativa;
- Espécies com capacidade de reprodução vegetativa: espécies que brotam e rebrotam, umas somente sob céu aberto e outras sob a sombra.

Os resultados de pesquisas a respeito de regeneração natural visam propor metodologias de manejo apropriadas para cada espécie, bem como de informações sobre o comportamento das espécies na floresta que, em consequência, servirão para uma infinidade de atividades relacionadas à reprodução das espécies nativas (SILVA; BRITEZ; SOUZA, 1993).

O objetivo do levantamento determina os limites da regeneração natural e constitui o apoio ecológico para a sobrevivência do ecossistema florestal (FINOL, 1969). Os indivíduos que constituem a regeneração natural são aqueles com DAP menor que 5 cm (ROLLET, 1969; VOLPATO, 1994). CARVALHO (1982) determinou como limite superior da regeneração o DAP de 15 cm e, segundo LONGHI (1980), pode-se considerar os indivíduos arbóreos com DAP menor que 20 cm inclusos nessa classificação.

Alguns autores consideram como regeneração natural todos os descendentes das plantas arbóreas que se encontram entre 0,1 m de altura até os 10 cm de diâmetro à altura do peito (DAP) (HOSOKAWA, 1986; FINOL, 1971). Essa será a definição adotada no presente estudo.

Mortalidade é o número de árvores que foram mensuradas, não foram cortadas e morreram durante o período de crescimento. Os fatores que causam a mortalidade são a idade ou senilidade, competição e supressão, doenças ou pragas, condições climáticas, fogos silvestres, anelamento e envenenamento, injúrias, corte ou abate da árvore (SANQUETTA, 1996).

A atividade meristemática gera o crescimento das árvores, ou seja, o alongamento e engrossamento das raízes, galhos e tronco, o aumento do peso, volume e forma da árvore.

Em uma floresta, o aumento em tamanho ocorrido em um intervalo de tempo é definido como crescimento e é dado pela atividade das árvores vivas, porém sua somatória não representa o crescimento da floresta como um todo, pois árvores morrem, são cortadas ou recrutadas. O crescimento das árvores, mais convenientemente medido pelo incremento da circunferência ou diâmetro à altura do peito é de grande interesse para a silvicultura e o manejo florestal (GOMIDE, 1997).

As variáveis utilizadas nessa pesquisa são o DAP (diâmetro a altura do peito), medido em centímetros, a uma altura de 1,30 metros, com precisão de 0,5 centímetros, utilizando-se fita métrica, e altura, medida em metros, utilizando-se fita métrica, com precisão de 0,1 metros.

Segundo ARCE et al. (1998), o crescimento das florestas pode ser entendido como um processo contínuo, que inclui uma entrada, um movimento e uma saída de matéria. A entrada é o ingresso, o movimento é o crescimento, e a saída é a mortalidade.

SASSAFRÁS

O sassafrás (*Ocotea odorifera* (Vellozo) J. G. Rohwer) é conhecido como canela-sassafrás, sassafrás amarelo, sassafrás rajado, sassafrás preto, sassafrás escuro em Santa Catarina e como canela em Minas Gerais (MATTOS FILHO, 1958). Essa espécie caracteriza-se por ser uma planta perenifólia, preferindo o alto das encostas de solos rasos e de rápida drenagem. Nas formações campestres de altitude, seu tronco é curto e a planta é mais baixa (10 m) com a copa bem arredondada. Produz anualmente irregular quantidade de sementes (LORENZI, 1998). Esta espécie apresenta crescimento monopodial, com os galhos distribuídos, na fase jovem, em pseudo-verticilos e inseridos em ângulos de 90° e sua região de ocorrência é do estado do Rio Grande do Sul até Minas Gerais. O sassafrás apresenta dispersão irregular e descontínua de sementes (INOUE; RODERJAN; KUNIYOSHI; 1984).

Segundo CARVALHO (2003) o sassafrás é uma planta esciófita e com crescimento relativamente lento quando jovem. Quanto ao grupo sucessional, o sassafrás é uma espécie secundária tardia ou clímax e é tolerante à sombra.

Na área de sua ocorrência natural, o sassafrás é uma espécie clímax, formadora de banco de mudas, produzindo sementes que germinam e sobrevivem sob a sombra (REITZ; KLEIN; REIS, 1978; REITZ; KLEIN; REIS, 1988).

TÓPICOS DE MATEMÁTICA

Há uma sensível e importante diferença entre os conceitos de “dado” e “informação”. O processamento de dados constitui uma ajuda porque reduz a quantidade de detalhes. Além disso, facilita a constatação de relações. “Dado” é o conjunto de valores medidos, mensurados ou observados, enquanto que seu processamento transforma-os em informações, organizando-os e condensando-os em gráficos ou em poucos números, os quais, então, transmitem a essência dos dados. O efeito consiste em eliminar detalhes menores e enfatizar os aspectos importantes dos dados (STEVENSON, 1981).

Segundo LEITHOLD (1982) variável real é um símbolo que representa qualquer um dos números de um conjunto real de números. Esse conjunto é o campo ou domínio da variável. Um constante é uma quantidade cujo valor permanece inalterado num dado problema. O campo ou domínio de uma constante é um número, simplesmente.

Entende-se por uma função f uma terna $(A, B, a \mapsto b)$ onde A e B são dois conjuntos e $a \mapsto b$, uma regra que permite associar a cada elemento a de A um único b de B . O conjunto A é o domínio de f e indica-se por D_f , assim $A=D_f$. O conjunto B é o contradomínio de f . O único b de B associado ao elemento a de A é indicado por $f(a)$ (GUIDORIZZI, 1987).

Seja S um espaço em que se defina uma função de probabilidade. Seja x uma função de valores reais definida em S (a função x transforma pontos de S em pontos de x). Diz-se que x é uma variável aleatória (variável aleatória unidimensional) (MOOD; GRAYBILL, 1978).

As variáveis aleatórias podem dividir-se em duas classes: 1) uma variável aleatória discreta é aquela que toma somente um número finito ou uma infinidade numerável de valores; 2) Uma variável aleatória contínua pode assumir qualquer valor de certo intervalo ou coleção de intervalos sobre esse x , sem a restrição de que aqueles sejam números isolados.

Uma variável discreta pode assumir um número finito ou infinito numerável de valores, enquanto que uma variável contínua pode assumir um número infinito de valores não numeráveis.

Diz-se que x é uma variável aleatória discreta unidimensional se é uma variável aleatória que toma somente um numero finito ou infinito numerável de valores desse x . Suponha-se que x toma unicamente valores $x_1, x_2, \dots, x_n, \dots$ com probabilidades $f(x_1), f(x_2), \dots, f(x_n), \dots$ e imagine-se que A é qualquer subconjunto dos pontos $x_1, x_2, \dots, x_n, \dots$, a probabilidade $P(A)$ do sucesso A (probabilidade de que x esteja em A) se define como $P(A) = \sum_A f(x)$, em que $\sum_A f(x)$ representa a soma de $f(x)$ para aqueles valores x_i que pertencem a A (MOOD; GRAYBILL, 1978).

MOOD; GRAYBILL (1978) define que x é uma variável aleatória contínua unidimensional se existe uma função f tal que $f(x) \geq 0$ para todo x do intervalo $-\infty < x < \infty$ e tal que para qualquer sucesso de A $P(A) = P(x \text{ esteja em } A) = \int_A f(x) dx$, em que $f(x)$ se denomina função densidade de x .

De acordo com FLEMMING; GONÇALVES (1992), sendo f uma função, o gráfico de f é o conjunto de todos os pontos $(x, f(x))$ de um plano coordenado, em que x pertence ao domínio de f .

Se uma função $f(x)$ é dada por $f(x) = c$, sendo c um número real, então diz-se que $f(x)$ é uma função constante (LEITHOLD, 1994).

Função identidade é a função $f: \mathbb{R} \rightarrow \mathbb{R}$ definida por $f(x) = x$. O gráfico dessa função é uma reta bissetriz dos quadrantes ímpares (FLEMMING, GONÇALVES, 1992).

As mesmas operações válidas para números são válidas para funções, ou seja, pode-se produzir novas funções efetuando-se as operações como adicionar, subtrair, multiplicar e dividir. FLEMMING; GONÇALVES (1992) afirmam que dadas as funções f e g , sua soma $f + g$, diferença $f - g$, produto $f \cdot g$ e quociente f/g , são definidas por:

$$a) (f + g)(x) = f(x) + g(x);$$

$$b) (f - g)(x) = f(x) - g(x);$$

$$c) (f \cdot g)(x) = f(x) \cdot g(x); \text{ e}$$

$$d) (f / g)(x) = \frac{f(x)}{g(x)}.$$

Se uma função for definida por $f(x) = a_n x^n + a_{n-1} x^{n-1} + \dots + a_1 x + a_0$, em que a_0, a_1, \dots, a_n são números reais $a_n \neq 0$ e n é um número inteiro não negativo, então $f(x)$ será chamada de função polinomial de grau n . Uma classificação mais abrangente define a função algébrica como aquela formada por um número finito de operações algébricas sobre as funções identidade e constante (SILVA, 2003).

Limites

Sejam f uma função e p um ponto do domínio de f ou extremidade de um dos intervalos que compõem o domínio de f . Diz-se que f tem limite L em p , se, para todo $\varepsilon > 0$ dado, existir um $\delta > 0$ tal que, para todo $x \in D_f$, $0 < |x - p| < \delta \Rightarrow |f(x) - L| < \varepsilon$. Tal número L , que quando existe é único, é indicado por $\lim_{x \rightarrow p} f(x)$. Assim,

$$\lim_{x \rightarrow p} f(x) = L \Leftrightarrow \begin{cases} \forall \varepsilon > 0, \exists \delta > 0 \text{ tal que, para todo } x \in D_f \\ 0 < |x - p| < \delta \Rightarrow |f(x) - L| < \varepsilon \end{cases} \quad (\text{GUIDORIZZI, 1987}).$$

De acordo com FLEMMING; GONÇALVES (1992), uma função f é contínua no ponto a se as seguintes condições forem satisfeitas:

- f é definida no ponto a ;
- $\lim_{x \rightarrow a} f(x)$ existe; e
- $\lim_{x \rightarrow a} f(x) = f(a)$.

Derivadas

A derivada de uma função $y = f(x)$ em relação a x é definida por $\frac{dy}{dx} = f'(x) = y' = \lim_{\Delta x \rightarrow 0} \frac{\Delta y}{\Delta x} = \lim_{\Delta x \rightarrow 0} \frac{f(x + \Delta x) - f(x)}{\Delta x}$ desde que o limite exista. A taxa instantânea de variação de uma função $y = f(x)$ em relação a x é dada por $\frac{dy}{dx}$ (PISKOUNOV, 1993).

A derivada fornece a inclinação da reta tangente à curva $y = f(x)$ no ponto $(x_1, f(x_1))$. Portanto, geometricamente, a derivada da função $y = f(x)$ no ponto x_1 representa a inclinação da curva nesse ponto (FLEMMING, GONÇALVES, 1992).

Toda função derivável num ponto x_1 é contínua nesse ponto (PISKOUNOV, 1993). Dado que $f(x)$ existe para todos os valores de $x \in (a, b)$ e f tem um extremo relativo em c , onde $a < c < b$, se $f'(c)$ existe, então $f'(c) = 0$.

Se a função f é derivável num intervalo aberto contendo c , se $(c, f(c))$ é um ponto de inflexão do gráfico de f , então se $f''(c)$ existe, $f''(c) = 0$ (LEITHOLD, 1982).

Integrais

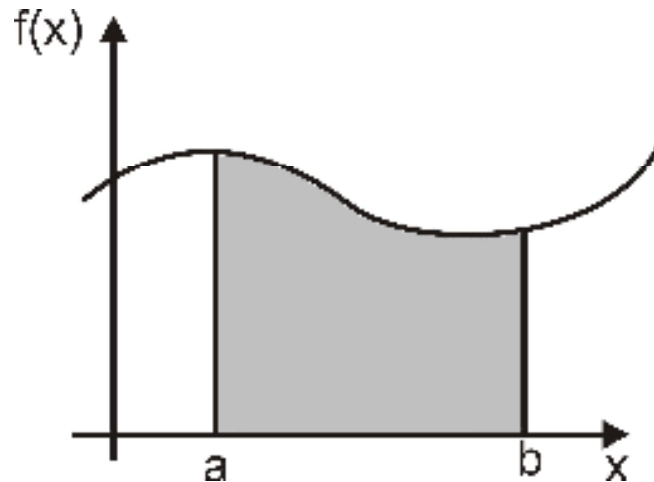
Teorema Fundamental do Cálculo

De acordo com SWOKOWSKI (1994), supondo-se f contínua em um intervalo fechado $[a, b]$:

- Se a função G é definida por $G(x) = \int_a^x f(t)dt$ para todo x em $[a, b]$, então G é uma antiderivada de f em $[a, b]$; e
- Se F é qualquer antiderivada de f em $[a, b]$, então $\int_a^b f(x)dx = F(b) - F(a)$.

FLEMMING; GONÇALVES (1992) citam que quando a função f é contínua e não negativa em $[a, b]$ a definição da integral definida coincide com a definição da área. Portanto, neste caso, a integral definida $\int_a^b f(x)dx$ é a área da região do gráfico delimitada entre a curva e o eixo das abscissas no intervalo de a até b .

FIGURA 1 – REPRESENTAÇÃO DA ÁREA DEFINIDA ENTRE A CURVA E O EIXO DAS ABCISSAS



Integrais impróprias

SWOKOWSKI (1994) define as integrais impróprias como:

- Se f é contínua em $[a, \infty)$ então $\int_a^\infty f(x)dx = \lim_{t \rightarrow \infty} \int_a^t f(x)dx$; e
- Se f é contínua em $(-\infty, a]$, então $\int_{-\infty}^a f(x)dx = \lim_{t \rightarrow -\infty} \int_t^a f(x)dx$, desde que o limite exista.

MODELAGEM

De acordo com STEVENSON (1981), um dos principais instrumentos extensamente usados na estatística é o modelo. Os modelos são versões simplificadas de algum problema ou situação da vida real. São usados para ilustrar certos aspetos da situação, evitando grande número de detalhes que talvez sejam irrelevantes para o problema; podem, assim, ajudar a reduzir o grau de complexidade de um problema.

Segundo MEYER (1969), todas as vezes que se emprega matemática a fim de estudar alguns fenômenos de observação deve-se essencialmente começar por construir um modelo matemático (determinístico ou probabilístico) para esses fenômenos. Inevitavelmente, o modelo deve simplificar as coisas e certos pormenores devem ser desprezados. O bom resultado do modelo depende de que os pormenores desprezados sejam ou não realmente sem

importância na elucidação do fenômeno estudado. A resolução do problema matemático pode estar correta e, não obstante, estar em grande discordância com os dados observados, simplesmente porque as hipóteses básicas feitas não sejam confirmadas. Geralmente é bastante difícil afirmar com certeza se um modelo matemático especificado é ou não adequado antes que alguns dados de observação sejam obtidos. A fim de verificar a validade de um modelo deve-se deduzir certo número de conseqüências do modelo e, a seguir, comparar esses resultados previstos com observações.

Um modelo determinístico é aquele que estipula condições sob as quais um experimento a ser executado determina o resultado do experimento. Em um modelo determinístico emprega-se “considerações físicas” para prever o resultado, enquanto que em um modelo não-determinístico (também chamado de probabilístico ou estocástico) emprega-se a mesma espécie de considerações para especificar uma distribuição de probabilidade (MEYER, 1969).

Para STEVENSON (1981),

a “estimação” é o processo que consiste em utilizar dados amostrais para estimar os valores de parâmetros populacionais desconhecidos. Essencialmente, qualquer característica de uma população pode ser estimada a partir de uma amostra aleatória. Entre os mais comuns estão a média e o desvio padrão de uma população e a proporção populacional. A avaliação de inventários, a estimação do custo de projetos, a avaliação de novas fontes de energia, predições sobre a realização de empreendimentos, estimativas de tempo médio são todas situações que envolvem estimação. A estimativa pontual é a estimativa única de um parâmetro populacional, enquanto que a estimativa por intervalo dá um intervalo de valores possíveis, no qual se admite esteja o parâmetro populacional.

A modelagem teve seu início no inventário florestal, uma vez que essa técnica possibilita a descrição quantitativa e qualitativa das florestas naturais ou artificiais. A descrição quantitativa é feita a partir do volume de madeira e do número de árvores disponíveis, enquanto que a descrição qualitativa das espécies contidas na floresta é feita considerando-se suas características e utilizações (COUTO s.d.)

O inventário fornece a situação da floresta no momento em que essa foi medida. (PENG, 2000).

A modelagem pode ser feita a partir de árvores individuais, que são obtidas pelo inventário de pontos permanentes e podem indicar a mortalidade do indivíduo em estudo, além de outros aspectos. A modelagem empírica é a base da modelagem utilizada em manejos florestais e a mecanicista é mais direcionada para pesquisas. A modelagem florestal está

intimamente ligada ao manejo de florestas, pois, com a condição de se predizer o futuro de uma floresta é mais fácil explorá-la racionalmente, sem que haja agressões ao meio ambiente (VANCLAY et al., 1997).

Para que um modelo matemático possa ser considerado uma função densidade de probabilidade, é necessário que as condições descritas no próximo item sejam atendidas.

TÓPICOS DE ESTATÍSTICA

A concepção profana de estatística usualmente inclui a coleção de grande quantidade de dados e a apresentação desses em tabelas ou gráficos; pode incluir também o cálculo de totais, médias, porcentagens, entre outros. Em todo caso, essas operações são apenas uma pequena parte da estatística (MOOD; GRAYBILL, 1978).

Para JOHNSON; LEONE (1964), a palavra estatística é usada em diversos contextos. Infelizmente algumas pessoas menos informadas a definem apenas como uma mera coleção de dados, tais como “A estatística do Brasil”, “O censo de um país”, “Tabelas de mortalidade”, “Estatística da produção de uma fazenda”, entre outros. Essas coleções de dados certamente são importantes fontes de informação, sendo pré-requisitos para um trabalho estatístico. Porém essa é apenas uma parte da estatística. A partir de coleta de dados sob condições controladas pode-se obter alguma informação que possibilite tirar conclusões.

MOOD; GRAYBILL (1978) salienta, sobre a origem da teoria estatística, que certas áreas da experimentação biológica alcançaram um ponto que, para que houvesse progresso, era imperativo o uso dos agora denominados métodos estatísticos. Foram então esses mesmos biólogos os que desenvolveram a parte essencial desta teoria. O curioso é que esse desenvolvimento deu-se de forma paralela a quase todos os campos do conhecimento abstrato.

Segundo JOHNSON; LEONE (1964), a estatística pode ser dividida em três partes: 1) Estatística descritiva: Compreende a organização, o resumo e, em geral, a simplificação de informações, que podem ser muito complexas; 2) Probabilidade: Útil para analisar situações que envolvam acaso, como distribuições diamétricas, volumétricas, falhas em mecanismos, jogos de dados e cartas; e 3) Inferência: Diz respeito à análise e interpretação de dados amostrais.

As origens da matemática da probabilidade remontam ao século XVI. As aplicações iniciais referiam-se quase todas a jogos de azar. Os jogadores ricos aplicavam o conhecimento da teoria das probabilidades para planejar estratégias de apostas. Mesmo hoje ainda há muitas aplicações que envolvem jogos de azar, tais como os diversos tipos de loteria, os cassinos, as corridas de cavalos e os esportes organizados. Todavia, a utilização das probabilidades

ultrapassou muito o âmbito desses jogos. Hoje os governos, as empresas, as organizações profissionais incorporam a teoria das probabilidades em seus processos diários de deliberações (STEVENSON, 1981).

Ainda de acordo com STEVENSON (1981), as probabilidades são utilizadas para exprimir a chance de ocorrência de determinado evento. A probabilidade de ocorrência de uma evento é dada por um número e que pode variar de 0 a 1.

O método clássico aplica-se à situações que tem resultados igualmente prováveis. Os jogos de azar usualmente apresentam esta característica de resultados igualmente prováveis. Calcular a probabilidade por esse método consiste em identificar o número de resultados “favoráveis” e dividi-lo pelo total de casos possíveis no espaço amostral. Segundo abordagem subjetiva, a probabilidade é uma avaliação pessoal do grau de viabilidade de um evento e está fortemente relacionada com a crença (STEVENSON, 1981).

A definição clássica de probabilidade, segundo MOOD; GRAYBILL (1978), diz que se um sucesso pode ocorrer de n maneiras mutuamente excludentes e igualmente verossímeis e se n_A dessas possuem um atributo A , a probabilidade de A é a fração $\frac{n_A}{n}$.

Embora nenhum plano de amostragem possa garantir que a amostra seja exatamente semelhante à população da qual foi extraída, uma amostra aleatória permite estimar o valor do erro possível, isto é, dizer quão próxima está a amostra da população, em termos de representatividade. As amostras não-aleatórias não apresentam estas características.

A parte do todo tomada é chamada amostra e o todo – do qual se extrai a amostra – é designado como população ou universo. Os elementos que compõem uma população podem ser indivíduos, firmas, produtos manufaturados, árvores ou parcelas, escolas, notas de aula, preços, ou qualquer coisa que possa ser mensurada, contada ou ordenada segundo postos (STEVENSON, 1981).

De acordo com STEVENSON (1981), o propósito da amostragem é fazer generalizações sobre a população básica; é axiomático que a população alvo seja estabelecida de modo que se possa fazer generalizações válidas. As populações limitadas em tamanho são denominadas finitas, enquanto que as não limitadas em tamanho são chamadas infinitas.

A amostragem é preferível ao censo quando:

- A população for infinita, pois o trabalho nunca acabaria;
- Se necessita informações em curto espaço de tempo;
- Se necessita destruir os itens para realizar o experimento (testes destrutivos);
- O custo do censo torna-o inviável economicamente; e

- A utilização de amostras permite mensurar variáveis que não seriam viáveis se fosse realizado um censo, em função do tempo e do custo. Assim, por meio da amostragem é possível realizar um estudo mais amplo e aprofundado.

O censo é preferível à amostragem quando:

- A população pode ser tão pequena que o custo e o tempo de um censo sejam pouco maior que para uma amostra;
- Se o tamanho da amostra é grande em relação ao da população, o esforço adicional requerido por um censo pode ser pequeno;
- Se é exigida a precisão completa, então o censo é o único procedimento aceitável; e
- Ocasionalmente já se dispõe de informação completa, de modo que não há necessidade de amostra.

A finalidade da amostragem, conforme descrito por JOHNSON; LEONE (1964), é obter uma indicação do valor de um ou mais parâmetros de uma população, tais como a média, o desvio padrão, ou a proporção de itens que possuem determinada característica. A estatística amostral que corresponde a esses parâmetros populacionais é usada para aproximar os valores desconhecidos daqueles parâmetros. Assim é que a média amostral é usada para estimar a média da população e o desvio padrão amostral é usado para estimar o desvio padrão populacional.

DISTRIBUIÇÕES

Uma distribuição de probabilidade ou de frequências relativas é a síntese do agrupamento de dados em um espaço amostral; mostra a proporção das vezes em que a variável aleatória tende a assumir cada um dos diversos valores (STEVENSON, 1981).

As distribuições descontínuas de probabilidades envolvem variáveis aleatórias relativas a dados que podem ser contados, como o número de ocorrências por amostra, ou o número de ocorrências por unidade num intervalo de tempo, de área, ou de distância (STEVENSON, 1981).

De acordo com MEYER (1969), se X uma variável discreta, portanto, R_x , o contradomínio de X , será formado no máximo por um número infinito numerável de valores x_1, x_2, \dots . A cada possível resultado x_i associa-se um número $p(x_i) = P(X = x_i)$, denominado probabilidade de x_i . Os números $p(x_i)$, $i = 1, 2, \dots$ devem satisfazer as seguintes condições, em que a função p é denominada função de probabilidade da variável aleatória X :

a) $p(x_i) \geq 0$, para todo i ,

b) $\sum_{i=1}^{\infty} p(x_i) = 1.$

Função densidade de probabilidade contínua, denotada por fdp, são limites de uma infinita seqüência de distribuições discretas, cujas variáveis formam um conjunto enumerável de valores. Tais distribuições devem verificar as seguintes condições, segundo SILVA, 2003:

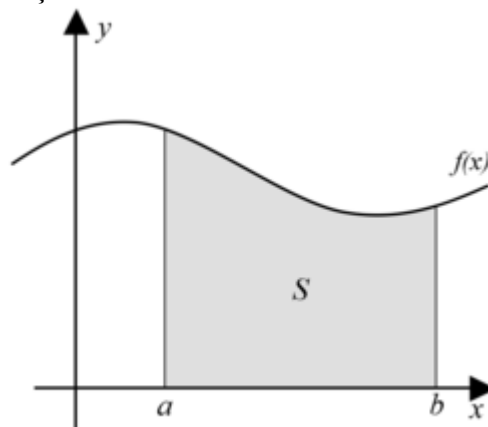
- $f(x) \geq 0$, para todos os valores de x , pois não existe probabilidade negativa;
- $\int_{-\infty}^{+\infty} f(x)dx = 1$, o que determina que a área entre a curva representativa da função $f(x)$ em todo o intervalo seja igual a 1; e
- $\int_a^b f(x)dx = P(a < X \leq b), b > a$, ou seja, a probabilidade da variável aleatória X

assumir valor em um intervalo será dada pela integral da função nesse intervalo. Geometricamente fica estabelecido que a probabilidade em um determinado intervalo corresponda à área determinada sob a região plana delimitada pela função nesse intervalo.

Propriedades decorrentes da definição de função densidade de probabilidade (MEYER, 1969):

- a) $P(a < X < b)$ representa a área sob a curva no gráfico da Figura 02, entre $x = a$ e $x = b$.

FIGURA 02 – REPRESENTAÇÃO DA ÁREA SOB A CURVA



- b) Uma consequência decorrente da descrição probabilística de X acima é que, para qualquer

valor especificado de X , x_0 , tem-se $P(X = x_0) = 0$, porque $P(X = x_0) = \int_{x_0}^{x_0} f(x)dx = 0$;

- c) Se uma função f satisfizer as condições $f(x) \geq 0$ para todo x , e $\int_{-\infty}^{+\infty} f(x)dx = k$, onde k é um número real positivo (não necessariamente igual a 1), então f não satisfaz todas as condições para ser uma fdp. No entanto, pode-se facilmente definir uma nova função g , em termos de f , assim, $g(x) = \frac{f(x)}{k}$ para todo x . Dessa maneira g satisfará todas as condições de uma fdp.
- d) Se X tomar valores somente em algum intervalo finito $[a, b]$, pode-se simplesmente arbitrar $f(x) = 0$ para todo $x \notin [a, b]$. Em conseqüência, a fdp ficará definida para todos os valores reais de x e pode-se exigir que $\int_{-\infty}^{+\infty} f(x)dx = 1$.

PRINCIPAIS DISTRIBUIÇÕES DISCRETAS

Segundo ROSS (1998), supondo um experimento em que o resultado possa ser classificado como sucesso ou falha e assumido $X = 1$ para sucesso e $X = 0$ para falha, então a função de probabilidade de X é dada por:

$$\begin{aligned} p(0) &= P\{X = 0\} = 1 - p \\ p(1) &= P\{X = 1\} = p \end{aligned}$$

Em que $0 \leq p \leq 1$ é a probabilidade de sucesso.

Esta variável aleatória X será chamada de variável aleatória de Bernoulli, para algum $p \in (0, 1)$.

O matemático suíço Jacques Bernoulli (1654 – 1705) foi o primeiro a estudar experimentos independentes, tendo uma probabilidade de sucesso comum p . Em seu livro *Ars Conjectandi* (a arte de conjecturar), publicado pelo seu sobrinho Nicolas oito anos após sua morte em 1713, Bernoulli mostrou que se um número de tais experimentos for grande, então a proporção deles que terá sucesso se aproximará de p com uma probabilidade próxima de 1. Jacques Bernoulli é da primeira geração da família dos mais famosos matemáticos de todos os tempos. Sua família fez contribuições fundamentais para a probabilidade, estatística e matemática. Uma dificuldade em saber o número exato de contribuições de cada membro da família consiste no fato de que alguns deles possuíam o mesmo nome. Outra dificuldade é que eles utilizavam nomes diferentes em diferentes localidades, por exemplo, Jacques, aqui citado, assinava também como Jaques. Entretanto, seus números, influência e resultados foram prodígios. Assim como os Bach para a música, os Bernoulli para a matemática constituíram uma família que entrou para a história (ROSS, 1998).

Usa-se o termo binomial para designar situações em que os resultados de uma variável aleatória podem ser agrupados em duas classes ou categorias. Os dados são, pois, nominais. As categorias devem ser mutuamente excludentes, de modo a deixar perfeitamente claro a qual categoria pertence determinada observação; e as classes devem ser coletivamente exaustivas, de forma que nenhum resultado fora delas é possível (STEVENSON, 1981). A distribuição binomial é, provavelmente, a de uso mais freqüente entre as distribuições discretas, nas aplicações da teoria estatística. Essa distribuição está associada com as provas repetidas de um mesmo sucesso (MOOD; GRAYBILL, 1978).

De acordo com STEVENSON (1981), a Distribuição de Poisson é útil para descrever as probabilidades do número de ocorrências num campo ou intervalo contínuo (em geral tempo ou espaço). Diz-se que a variável x se distribui segundo uma distribuição de Poisson se a função de quantidade é $f(x) = \frac{e^{-m} m^x}{x!}$, com $x = 0, 1, 2, \dots$ em que m é qualquer número positivo (MOOD; GRAYBILL, 1978). Essa distribuição baseia-se nas seguintes hipóteses:

1. A probabilidade de uma ocorrência é a mesma em todo o campo de observação;
2. A probabilidade de mais de uma ocorrência num único ponto é aproximadamente zero; e
3. O número de ocorrências em qualquer intervalo é independente do número de ocorrências em outros intervalos.

A Distribuição Multinomial, segundo STEVENSON (1981), é usada em situações onde há mais de dois resultados mutuamente excludentes. Tal como na binomial, exige-se que as provas sejam independentes, com probabilidade constante. Um exemplo de distribuição multinomial é a jogada de um dado.

A Distribuição Hipergeométrica refere-se a situações com dois ou mais resultados, em que a probabilidade de sucesso varia de uma prova para outra (STEVENSON, 1981). A função de quantidade hipergeométrica é

$$f(x) = \frac{\binom{m}{x} \binom{n}{r-x}}{\binom{m+n}{r}}, \text{ com } x = 0, 1, 2, \dots, r \text{ (MOOD; GRAYBILL, 1978).}$$

PRINCIPAIS DISTRIBUIÇÕES CONTÍNUAS

Dentre as Funções Densidade de Probabilidade clássicas estão a Exponencial, Gamma, Beta, Weibull e Normal (Gauss), que serão apresentadas a seguir.

DISTRIBUIÇÃO EXPONENCIAL

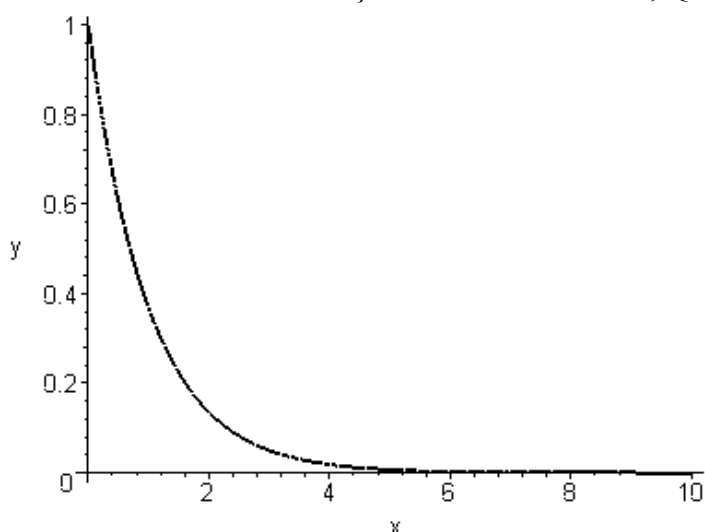
A Distribuição Exponencial desempenha um papel importante na descrição de uma grande classe de fenômenos, particularmente nos assuntos da teoria da confiabilidade (MEYER, 1974). Para JOHNSON; LEONE (1964), esse modelo pode ser considerado como um caso particular da Distribuição Gamma, que será apresentada no tópico a seguir. A distribuição exponencial tem sido utilizada para prever o período de tempo necessário até um evento ocorrer, por exemplo o tempo até acontecer um terremoto, ou até começar uma nova guerra ou até o telefone tocar, você atender e perceber que discaram errado ou até quando vai uma lâmpada expirar ou queimar (ROSS, 1998). A distribuição exponencial, de acordo com STEVENSON (1981), envolve probabilidade ao longo do tempo ou da distância entre ocorrências num intervalo contínuo. Por exemplo, a exponencial é usada como modelo do tempo entre falhas de equipamento elétrico, tempo de chegada de clientes em um supermercado, tempo entre chamadas telefônicas, entre outros. No âmbito das Ciências Florestais a Distribuição Exponencial é usada como modelo para avaliação de regeneração natural.

Ainda segundo MEYER (1974), a função densidade de probabilidade exponencial, com variável aleatória contínua X , é dada pela Equação 01:

$$f(x) = \begin{cases} \frac{1}{\beta} \cdot e^{\frac{-x}{\beta}} & \text{para } x \geq 0, \beta > 0 \\ 0 & \text{e.o.c.} \end{cases} \quad \text{EQUAÇÃO 01}$$

O gráfico do modelo é mostrado na Figura 03

FIGURA 03 – DISTRIBUIÇÃO EXPONENCIAL, QUANDO $\beta = 1$



A média e a variância, através da expectativa matemática, são calculadas através das seguintes fórmulas, respectivamente, (EQUAÇÕES 02 e 03):

$$E(x) = \beta^{-1} \quad \text{EQUAÇÃO 02}$$

$$\sigma_x^2 = \beta^{-2} \quad \text{EQUAÇÃO 03}$$

Esta distribuição intercepta o eixo y no ponto $\frac{1}{\beta}$ pois:

- Tomando-se a Distribuição Exponencial:

$$y = \frac{1}{\beta} \cdot e^{\frac{-x}{\beta}}$$

- Substituindo $x = 0$

$$y = \frac{1}{\beta} \cdot e^{\frac{-0}{\beta}}$$

$$y = \frac{1}{\beta} \cdot e^0$$

- Como:

$$e^0 = 1$$

- Logo:

$$y = \frac{1}{\beta} \text{ c.q.d}$$

DISTRIBUIÇÃO GAMMA

A Distribuição Gamma é definida a seguir (EQUAÇÃO 04):

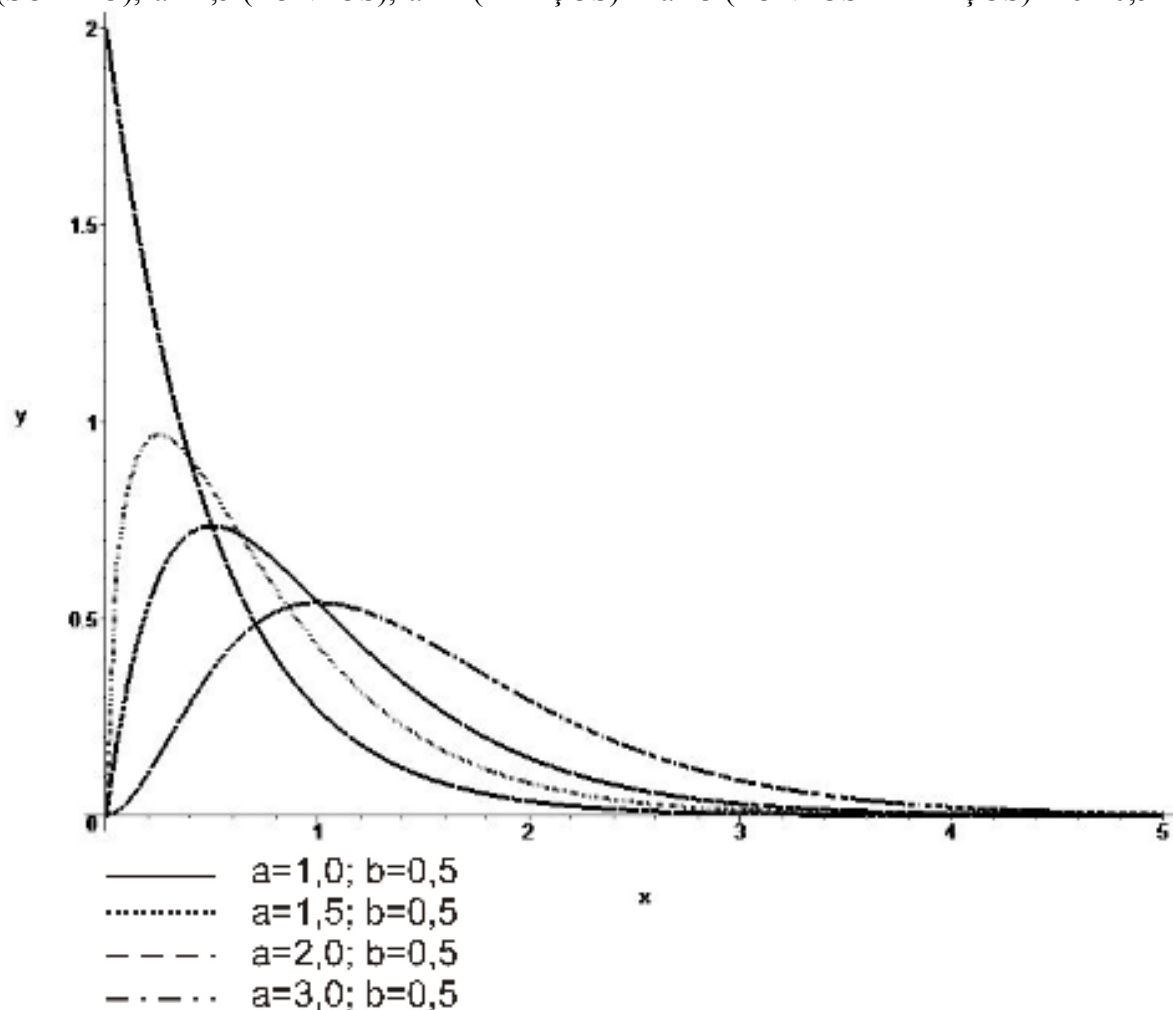
$$f(x) = \begin{cases} \frac{x^{a-1} \cdot e^{-x/b}}{b^a \cdot \Gamma(a)} & \text{para } x \geq 0 \\ 0 & \text{e.o.c,} \end{cases} \quad \text{EQUAÇÃO 04}$$

Em que a e b são valores positivos e afetam a forma e a escala, respectivamente (ROUSSAS, 1997).

A Figura 04 representa a curva gerada a partir da Função Gamma, em que $b = 0,5$ e a varia entre 1 e 3.

A Distribuição Gamma com $a = \frac{1}{2}$ e $b = \frac{n}{2}$ (n sendo um inteiro positivo) é chamada de distribuição χ_n^2 (lê-se qui-quadrado) com n graus de liberdade.

FIGURA 04 - REPRESENTAÇÃO GRÁFICA DA DISTRIBUIÇÃO GAMMA, PARA $a=1$ (SÓLIDO), $a=1,5$ (PONTOS), $a=2$ (TRAÇOS) E $a=3$ (PONTOS E TRAÇOS) E $b=0,5$



A média e a variância são dadas pelas equações 05 e 06, respectivamente.

$$E(x) = a \cdot b \quad \text{EQUAÇÃO 05}$$

$$\sigma^2 = a \cdot b^2 \quad \text{EQUAÇÃO 06}$$

DISTRIBUIÇÃO BETA

Considera-se a Função Beta como sendo (EQUAÇÃO 07):

$$f(x) = \begin{cases} \frac{\Gamma(a+b)}{\Gamma(a) \cdot \Gamma(b) \cdot (c-d)^{a+b-1}} \cdot (x-c)^{a-1} \cdot (d-x)^{b-1} & \text{se } c < x < d \\ 0 & \text{e.o.c.} \end{cases} \quad \text{EQUAÇÃO 07}$$

A Distribuição Beta pode assumir diversas formas, de acordo com os valores de a e b , como observa-se nas Figuras 05 a 08.

A Beta pode ser usada como modelo de fenômenos aleatórios cujo conjunto de valores possíveis é um intervalo finito $[c,d]$, ou seja, no intervalo de c até d , c como unidade de medida pode ser transformada para um intervalo $[0,1]$.

FIGURA 05 – REPRESENTAÇÃO GRÁFICA DA DISTRIBUIÇÃO BETA, PARA $a=1$ E $b=2$ (SÓLIDO), $a=2$ E $b=1$ (PONTOS) E $a=1$ E $b=1$ (TRAÇOS)

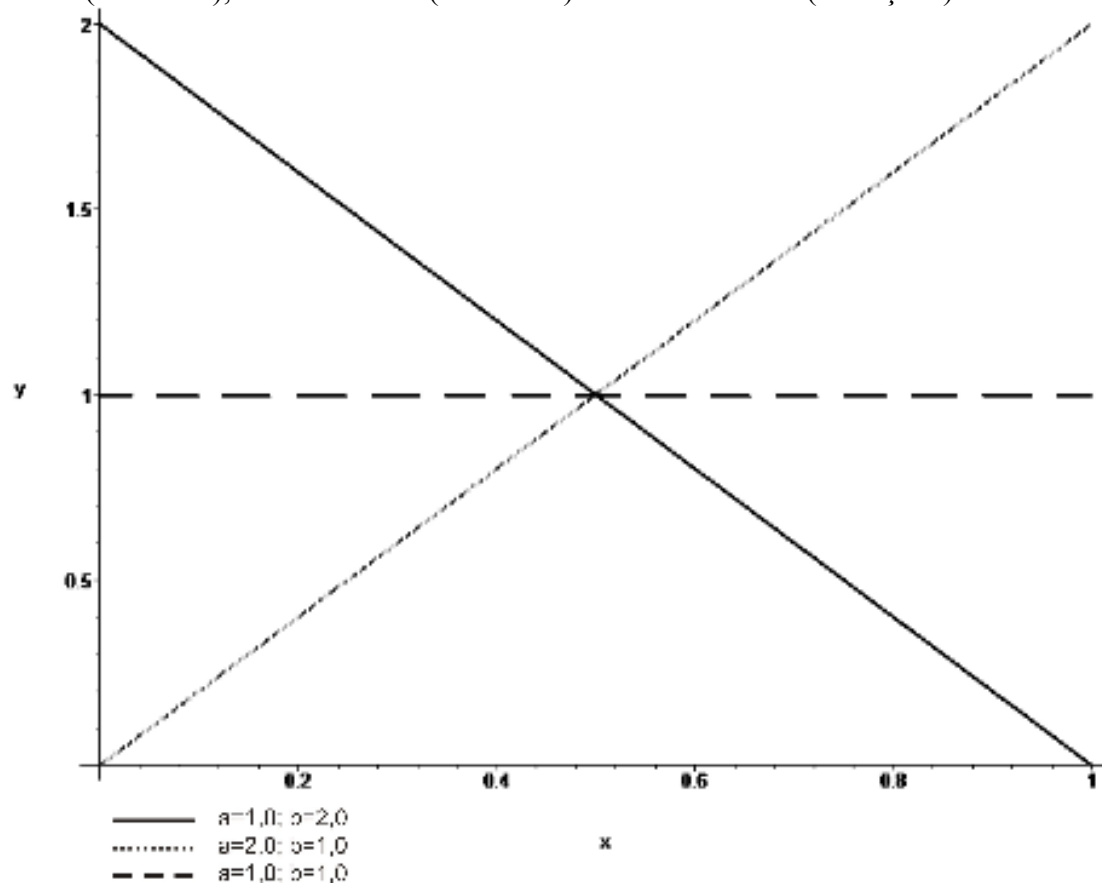


FIGURA 06 – REPRESENTAÇÃO GRÁFICA DA DISTRIBUIÇÃO BETA, PARA $a=0,3$ E $b = 0,8$ (SÓLIDO) E $a=0,3$ E $b = 0,3$ (PONTOS)

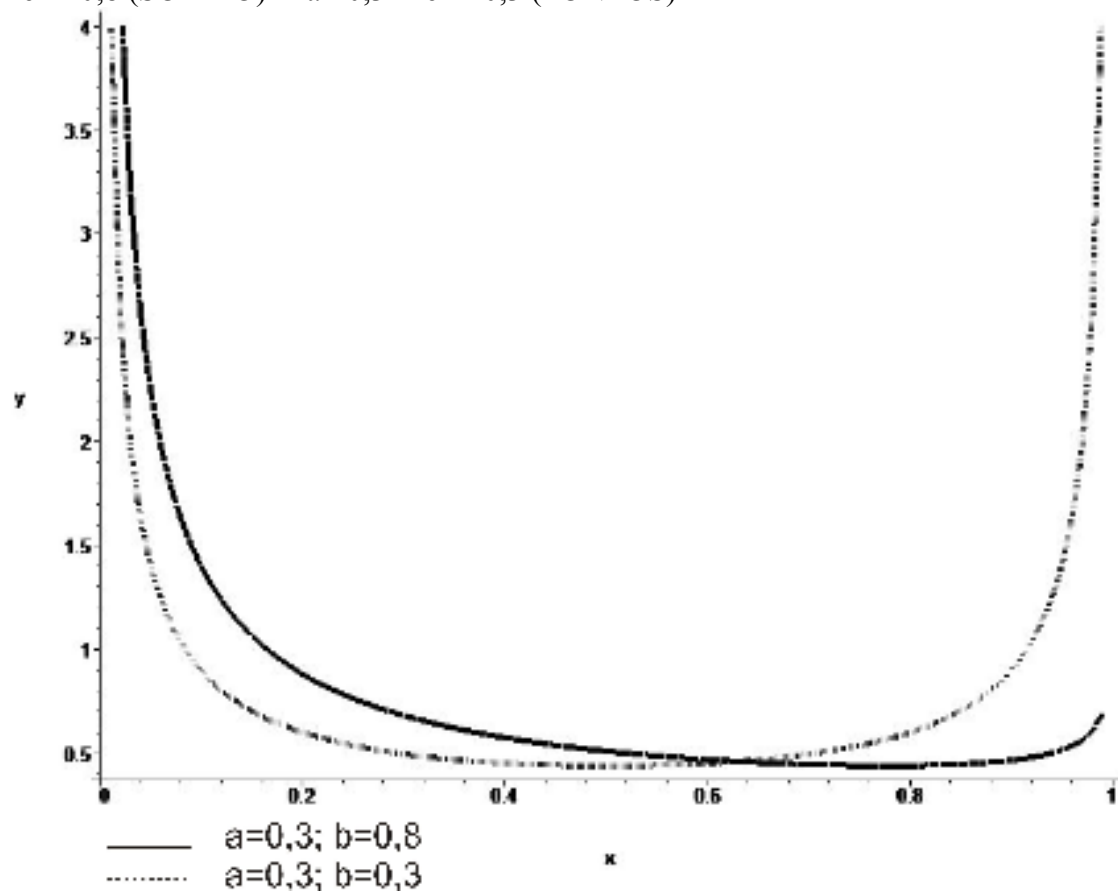


FIGURA 07 – REPRESENTAÇÃO GRÁFICA DA DISTRIBUIÇÃO BETA, PARA $a=2$ E $b = 3$ (SÓLIDO), $a=5$ E $b = 3$ (PONTOS) E $a=2$ E $b = 2$ (TRAÇOS)

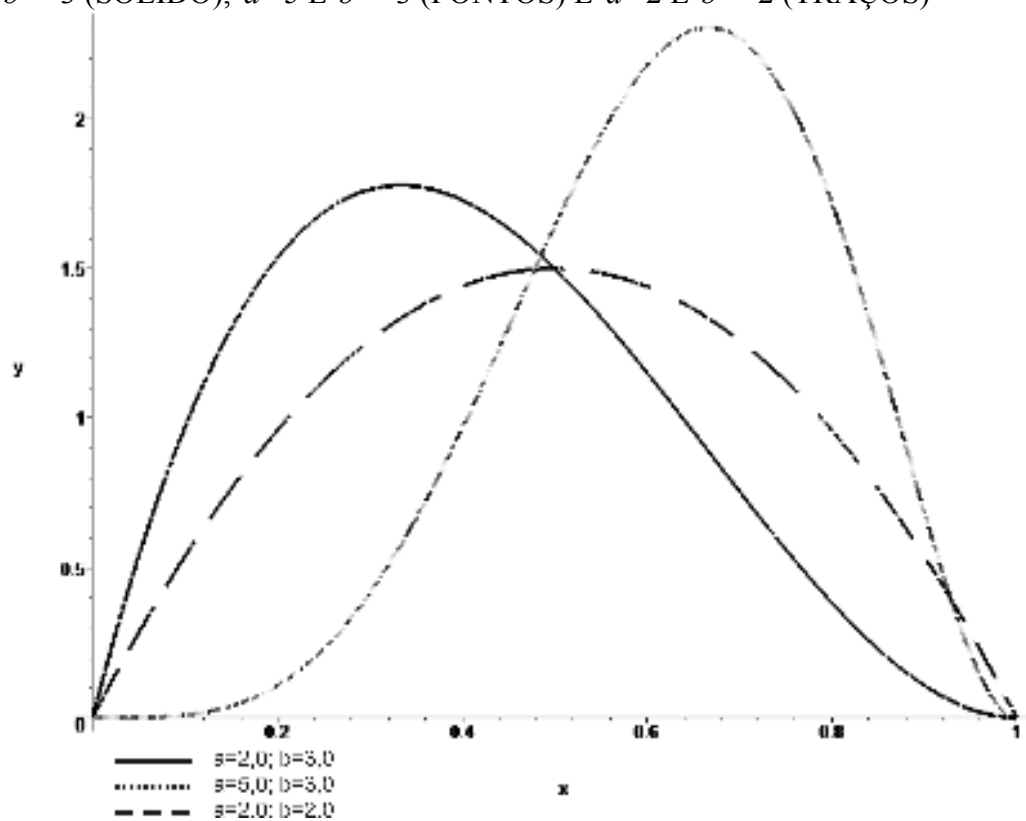
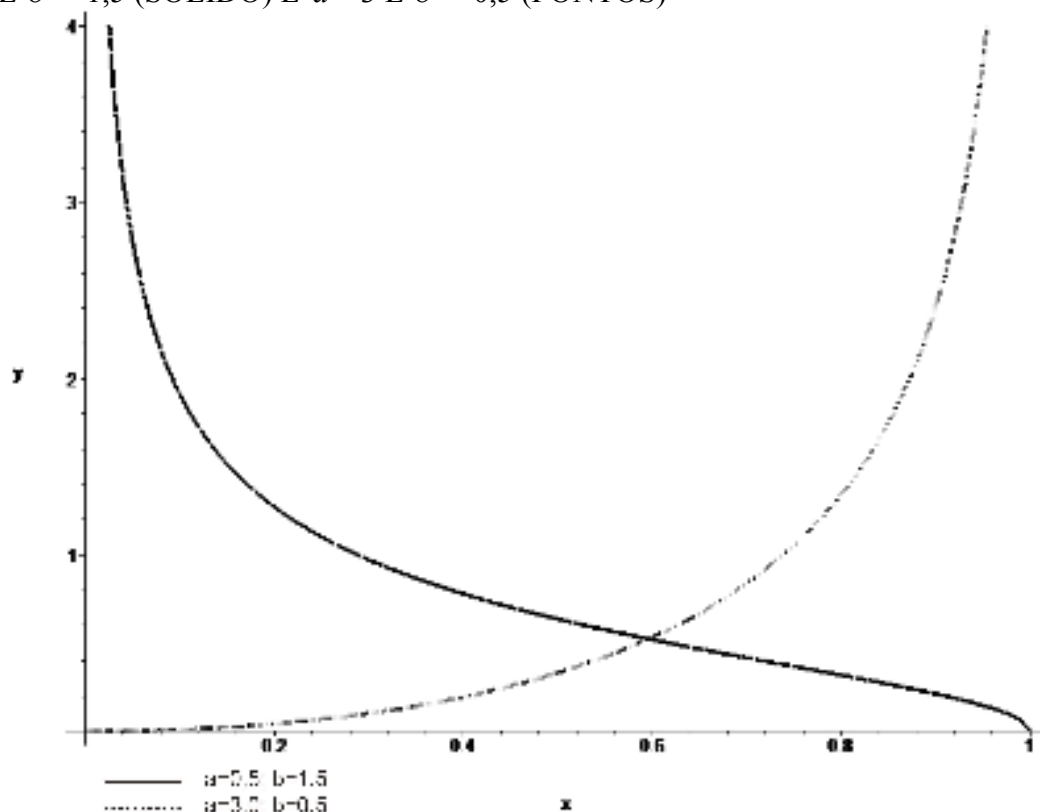


FIGURA 08 – REPRESENTAÇÃO GRÁFICA DA DISTRIBUIÇÃO BETA, PARA $a = 0,5$ E $b = 1,5$ (SÓLIDO) E $a = 3$ E $b = 0,5$ (PONTOS)



A média é dada pela Equação 08 e a variância pela Equação 09.

$$E(x) = \frac{a}{a+b} \quad \text{EQUAÇÃO 08}$$

$$\sigma^2 = \frac{a \cdot b}{(a+b)^2 \cdot (a+b+1)} \quad \text{EQUAÇÃO 09}$$

DISTRIBUIÇÃO WEIBULL

Esta distribuição é amplamente utilizada para testes de confiabilidade e vida útil (JOHNSON; LEONE, 1964). A função densidade de probabilidade da Distribuição WEIBULL apresenta-se conforme a seguinte expressão (EQUAÇÃO 10):

$$f(x) = \begin{cases} \frac{c}{b} \cdot \left(\frac{x-a}{b}\right)^{c-1} \cdot e^{-\left(\frac{x-a}{b}\right)^c} & \text{se } x \geq a, b > 0 \text{ e } c > 0 \\ 0 & \text{e.o.c.} \end{cases} \quad \text{EQUAÇÃO 10}$$

Assim como a função beta, a Weibull também pode assumir várias formas (FIGURA 09).

A Distribuição Weibull é usualmente utilizada na engenharia devido à sua versatilidade. Foi originalmente proposta para interpretação de fadiga de materiais em Stockholm por Ernst Hjalmar Waloddi Weibull em 1939, mas atualmente seu uso estendeu-se a muitos outros problemas de engenharia.

Em particular, é comumente usada no campo dos fenômenos da vida, como uma distribuição de tempo de vida de algum objeto particularmente quando o modelo de ligação mais fraca é apropriado para o objeto. Isto é, considerar um objeto com muitas partes e supor que existirá uma falha no objeto. Sob essas condições foi mostrado que a Distribuição Weibull fornece uma boa aproximação com a distribuição do tempo de vida do item (ROSS, 1998).

A média e a variância são dadas pelas Equações 11 e 12 respectivamente.

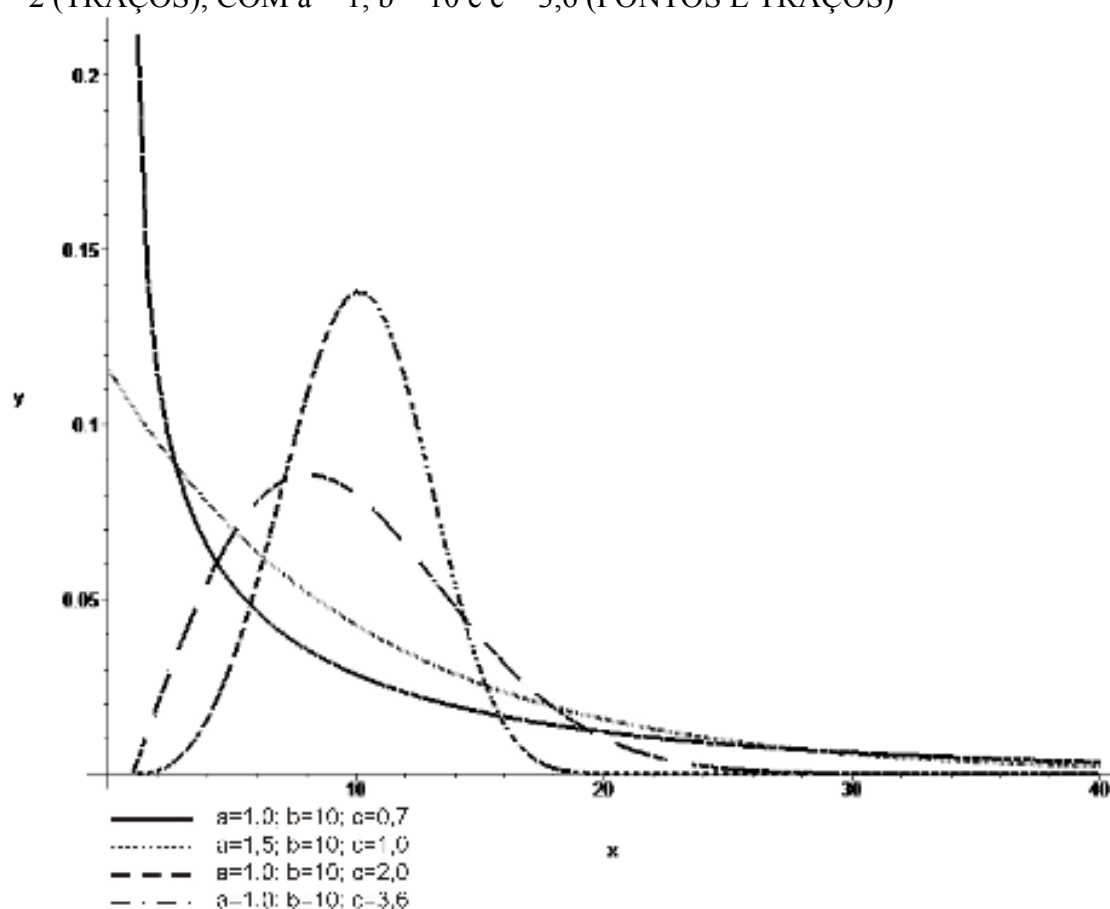
$$E(x) = a + b \cdot \Gamma\left(1 + \frac{1}{c}\right) \quad \text{EQUAÇÃO 11}$$

$$\sigma_x^2 = b^2 \cdot \sigma_y^2 \quad \text{EQUAÇÃO 12}$$

em que σ_y^2 é calculado pela Equação 13.

$$\sigma_y^2 = \Gamma\left(1 + \frac{2}{c}\right) - \Gamma^2\left(1 + \frac{1}{c}\right) \quad \text{EQUAÇÃO 13}$$

FIGURA 09 - REPRESENTAÇÃO GRÁFICA DA DISTRIBUIÇÃO WEIBULL, COM $a = 1$, $b = 10$ e $c = 0,7$ (SÓLIDO), COM $a = 1$, $b = 10$ e $c = 1$ (PONTOS), COM $a = 1$, $b = 10$ e $c = 2$ (TRAÇOS), COM $a = 1$, $b = 10$ e $c = 3,6$ (PONTOS E TRAÇOS)



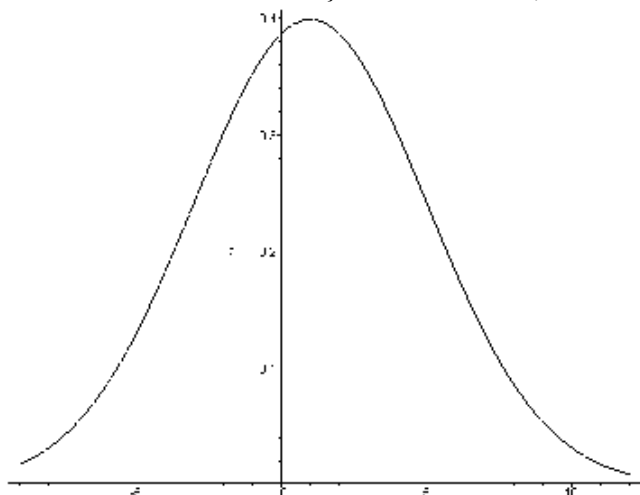
DISTRIBUIÇÃO NORMAL OU DE GAUSS

De acordo com MEYER (1974), a Distribuição Normal, desenvolvida por De Moivre, é de extrema importância no campo das probabilidades, com aplicações em diversas áreas do conhecimento. Esse modelo é definido como se segue (EQUAÇÃO 14):

$$f(x) = \begin{cases} \frac{1}{\sqrt{2\pi} \cdot \sigma} \cdot e^{-\frac{1}{2} \left(\frac{x-\mu}{\sigma} \right)^2} & -\infty < x < +\infty \\ 0 & \text{e.o.c.} \end{cases} \quad \text{EQUAÇÃO 14}$$

A Distribuição Normal pode ser representada graficamente pela Figura 10:

FIGURA 10 – DISTRIBUIÇÃO NORMAL, $\sigma = 1$ E $\mu = 0$



As Equações 15 e 16 representam a média e a variância, respectivamente.

$$E(x) = \text{média aritmética} = \mu \quad \text{EQUAÇÃO 15}$$

$$\text{var}(x) = \sigma^2 \quad \text{EQUAÇÃO 16}$$

Esta distribuição possui as seguintes propriedades:

- A curva normal tem forma de sino;
- É simétrica em relação à média;
- Prolonga-se de $-\infty$ a $+\infty$;

- Cada distribuição normal fica completamente especificada por sua média e seu desvio padrão; há uma distribuição normal distinta para cada combinação de média e desvio padrão;
- A área total sob a curva é considerada como 100%;
- A área sob a curva entre dois pontos é a probabilidade de uma variável normalmente distribuída tomar um valor entre esses pontos;
- Como há um número ilimitado de valores no intervalo $-\infty$ a $+\infty$, a probabilidade de uma variável aleatória distribuída normalmente tomar exatamente determinado valor é zero. Assim, as probabilidades se referem a intervalos de valores; e
- A área sob curva entre a média e um ponto arbitrário é função do número de desvios padrões entre a média e aquele ponto.

A velocidade em qualquer direção, o erro na medida de uma quantidade física, a distribuição de frequência de volumes e diâmetros em uma floresta, possuem um comportamento que aproxima-se da distribuição de probabilidades normal (ROSS, 1998).

A distribuição normal foi introduzida pelo matemático francês Abraham De Moivre, em 1733. De Moivre, que usou essa distribuição para aproximar probabilidades referentes a lançamentos de moedas, a chamou de curva exponencial em forma de sino (exponential bell-shaped curve). Entretanto, foi apresentada realmente apenas em 1809, quando o famoso matemático alemão K. F. Gauss a usou como uma parte integral da sua aproximação para prever o local de entidades astronômicas. Como resultado tornou-se comum após essa data chamá-la de distribuição Gaussiana (ROSS, 1998).

Do meio para o final do século XIX, entretanto, muitos estatísticos passaram a acreditar que a maioria dos intervalos de dados teria histogramas de acordo com a forma de sino Gaussiano. Além disso, tornou-se aceitável que fosse “normal” para qualquer conjunto de dados bem comportados seguir esta curva. Como resultado, seguindo o estatístico britânico Karl Pearson, as pessoas começaram a se referir à curva de Gauss chamando-a de curva normal (ROSS, 1998).

Hodiernamente não faltam consultores em estatística, muitos trabalhando em escritórios elegantes. Entretanto, no começo do século XIX, o primeiro nesse ramo de trabalho ganhava dinheiro no escuro, apostando em casas de jogos em Long Acres, Londres, conhecidas como casas de café. Ele era Abraham De Moivre, um protestante refugiado da França Católica e, por um preço, ele poderia computar a probabilidade de ganhar em jogos de azar em todos os tipos de jogos (ROSS, 1998).

Entretanto, De Moivre, o descobridor da curva normal, fez sua vida em uma loja de café, ele era um matemático de habilidades reconhecidas. Além disso, ele era um membro da sociedade real e foi convocado por Sir Isaac Newton. Escutando a Karl Pearson e imaginando De Moivre na casa de

café: “Eu pinto De Moivre trabalhando em uma mesa suja na casa de café com um jogador quebrado ao seu lado e Isaac Newton andando pelo corredor observando seu amigo. Daria um grande quadro para um artista inspirado” (ROSS, 1998).

Karl Friederich Gauss (1777 – 1855), um dos primeiros usuários da curva normal, foi um dos maiores matemáticos de todos os tempos. O conhecido matemático historiador E. T. Bell expressou em seu livro “Men of mathematics” (Homens da matemática), em um capítulo intitulado “The prince of mathematicians” (O príncipe dos matemáticos): “Arquimedes, Newton e Gauss; esses três estão em uma classe por si mesmos entre grandes matemáticos e não seriam pobres mortais capazes de classificá-los em ordem de mérito. Os três iniciaram tendências em matemática pura e aplicada. Arquimedes estimou sua matemática pura mais que suas aplicações; Newton pareceu ter encontrado a justificativa chefe para suas invenções matemáticas nos usos científicos nos quais ele colocou; enquanto Gauss declarou que para tudo trabalha-se no lado puro e aplicado” (ROSS, 1998).

3 - METODOLOGIA

LOCALIZAÇÃO DA ÁREA

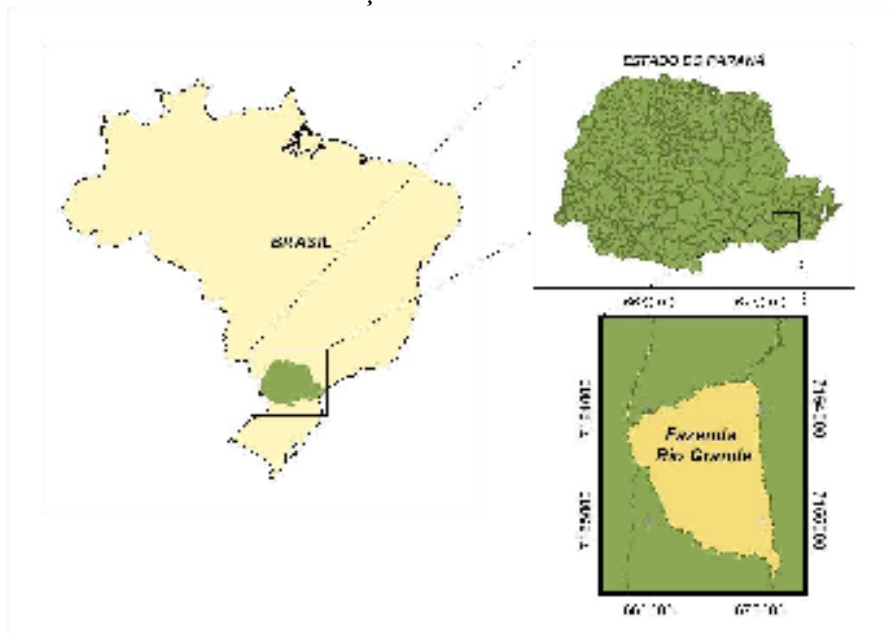
A Fazenda Experimental Gralha Azul localiza-se no município de Fazenda Rio Grande, Região Metropolitana de Curitiba, no estado do Paraná, Primeiro Planalto Paranaense (FIGURA 11; 12). Essa unidade de pesquisa está situada numa área de aproximadamente 877 hectares, onde estão instaladas 6 (seis) parcelas permanentes do Programa Ecológico de Longa Duração (PELD), SITE 09, onde foram coletados dados de DAP (cm) (diâmetro a altura do peito) e altura (m) de regenerações naturais e árvores adultas de Sassafrás (*Ocotea odorifera*).

Neste centro de pesquisas estão instaladas avicultura de corte, avicultura de postura, cunicultura, bovinocultura de leite, ovinocultura, hospital veterinário, fruticultura, pastagens, lavouras anuais, suinocultura, bovinocultura de leite e garagem de máquinas.

Esta unidade situa-se entre as seguintes coordenadas geográficas: Latitude 25°37'32" S e 25°41'33" S; Longitude 49°15'29" W e 49°17'27" W, possui altitudes entre 870 e 920 m e pertence à Bacia Hidrográfica do Alto Iguaçu.

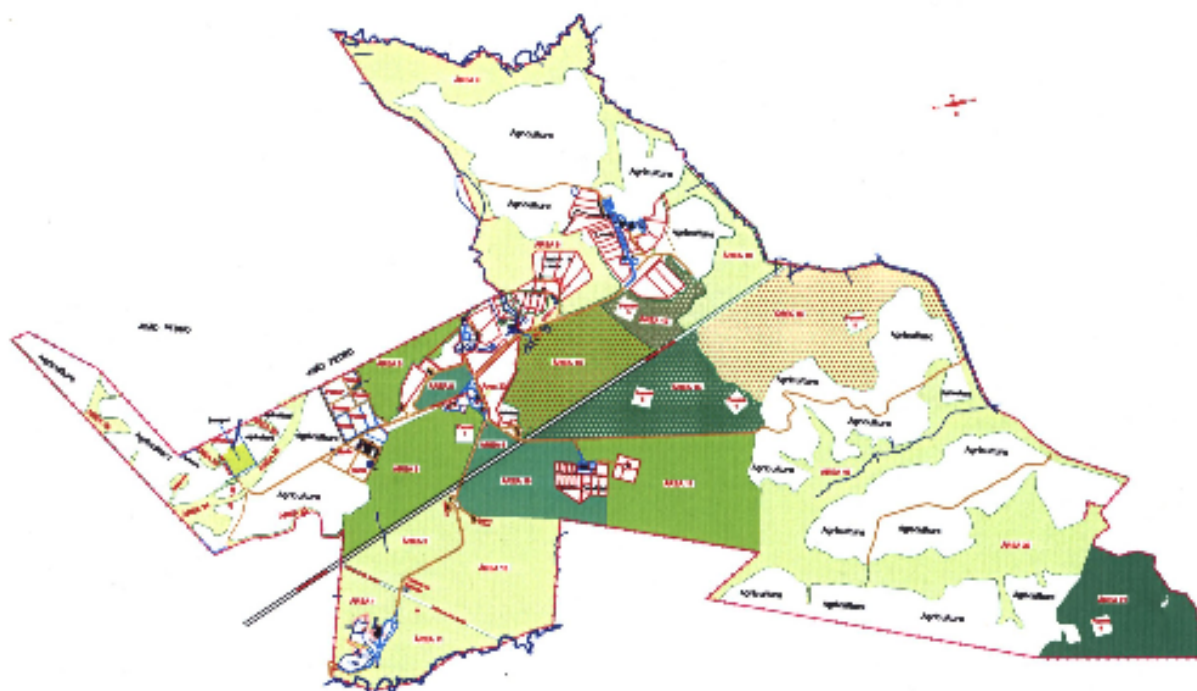
O Acesso principal dá-se pela Rodovia BR 116 no quilômetro 21,5.

FIGURA 11 – LOCALIZAÇÃO DA FAZENDA EXPERIMENTAL GRALHA AZUL



Fonte: PUCPR

FIGURA 12 – MAPA DA ÁREA DA UNIDADE EXPERIMENTAL



Fonte: RELATÓRIO SASSAFRÁS

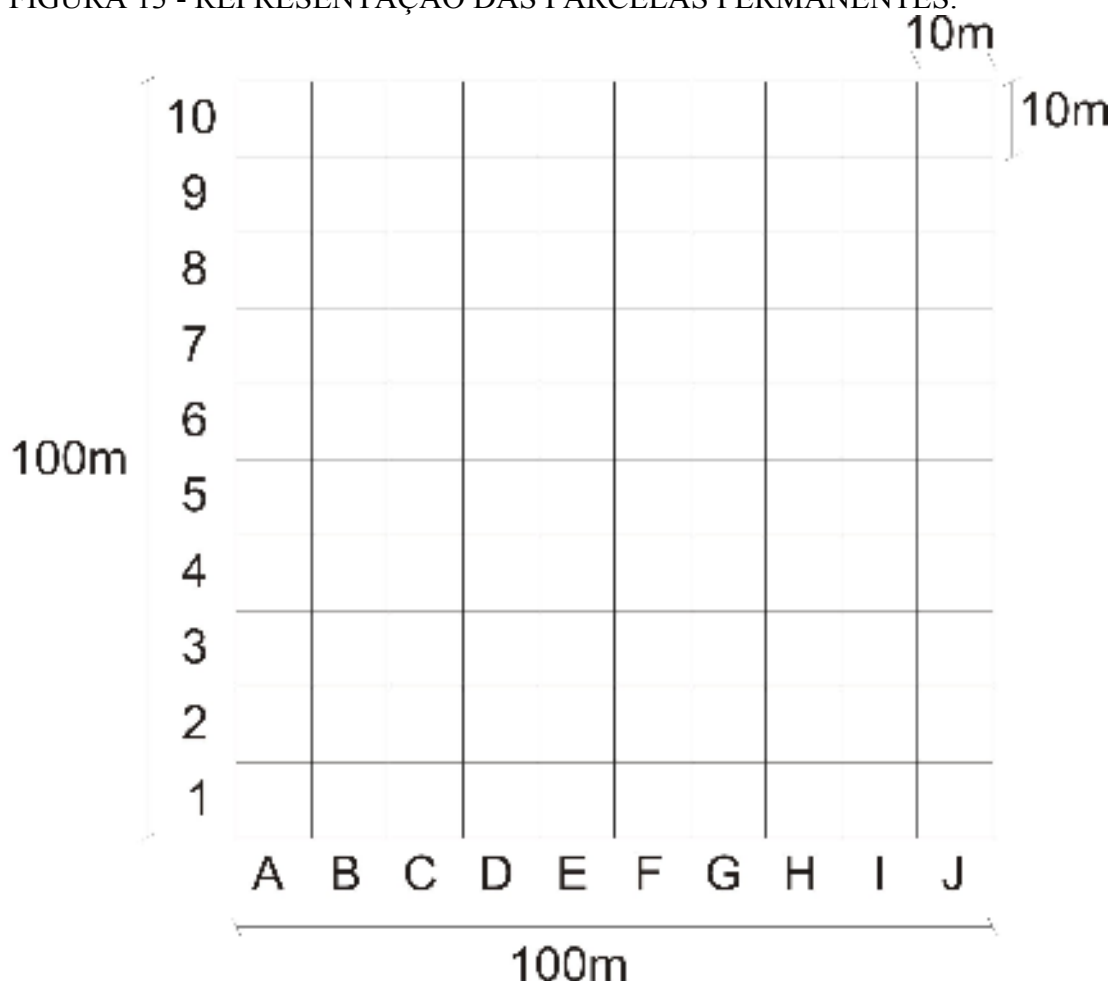
COLETA DE DADOS

Nas áreas em que se realizou o presente estudo foram encontradas até 120 espécies arbóreas por parcela, o que aumenta a complexidade de estudar-se sua regeneração natural.

O inventário detalhado da regeneração natural do Sassafrás: *Ocotea odorifera*, espécie da família *Lauraceae*, foi realizado definindo-se seis parcelas permanentes de 1 hectare e subdividindo-as em 100 sub-unidades de 100 m² cada (PROJETO SASSAFRÁS: Bioecologia e Uso Sustentável).

Dividiu-se cada parcela em dez fileiras, nomeadas pelas letras A, B, C, D, E, F, G, H, I e J e, cada fileira em dez blocos (1, 2, 3, 4, 5, 6, 7, 8, 9, 10), de acordo com a figura 13. Procedeu-se à identificação e medição das regenerações naturais pela Fileira A, prosseguindo-se em zig-zag pelas demais fileiras até a Fileira J (PELD – Programa Ecológico de Longa Duração).

FIGURA 13 - REPRESENTAÇÃO DAS PARCELAS PERMANENTES.



Considerou-se como regeneração natural todas as árvores com DAP (Diâmetro à Altura do Peito) menor ou igual a 10 cm (dez centímetros), em que se mediu o diâmetro à altura do colo e altura. A distribuição de alturas foi utilizada para o presente estudo, considerando-se intervalos de classe de 1 m (um metro).

Os bolsistas do Projeto Sassafrás mediram cem por cento das mudas de sassafrás classificadas como regeneração natural das seis parcelas permanentes. Todas as regenerações naturais das seis parcelas permanentes foram localizadas e identificadas através de marcação com etiquetas vermelhas, numeradas sequencialmente, precedidas da letra S (Sassafrás), além do referenciamento adotado para a localização espacial nas parcelas das regenerações através de um sistema de coordenadas (x, y), cujos valores de x e de y foram tomados de zero a dez metros, a partir de cada sub-unidade ou bloco (CALIJURI, M. L.; RÖHM, 1995).

Salienta-se que os modelos foram ajustados à distribuição de alturas da regeneração natural do Sassafrás, pois se considera, por definição teórica, que não haja indivíduos com altura 0 (zero).

AJUSTE DOS MODELOS

As pesquisas realizadas no Projeto Sassafrás revelaram que a distribuição de alturas, da regeneração natural da espécie arbórea Sassafrás é decrescente. Essa situação é representada usualmente pelo modelo exponencial, mas outras funções podem também ser ajustadas a esse tipo de dados. Para a presente pesquisa foram ajustadas as Distribuições de Probabilidade Exponencial, Gamma, Beta, Weibull e Normal, em que se utilizou a distribuição de alturas da regeneração natural.

Foram escolhidas 5 (cinco) funções matemáticas para serem ajustadas, com o auxílio do software *Table Curve 2d*, aos dados de regeneração natural e a que resultou em melhor aderência foi escolhida. Essa função foi transformada em função densidade de probabilidade, verificando-se as premissas descritas anteriormente, e foram desenvolvidas as expressões que, por meio dos coeficientes, através das quais são calculadas a média e a variância do modelo.

Para efetuar-se a comparação entre os modelos, foram calculados o coeficiente de determinação múltiplo, o erro padrão da estimativa recalculado em percentagem, teste do χ^2 e o de Kolmogorov-Smirnov. Procedeu-se uma verificação qualitativa dos resultados, a partir da análise de resíduos, a fim de se verificar se os modelos são tendenciosos ou possuem variância homogênea ao longo de toda a distribuição.

Há, no campo da matemática, duas maneiras de se ajustar um modelo: por meio de regressão linear ou não linear.

A regressão linear caracteriza-se basicamente por ajustar parâmetros de funções polinomiais de grau n , sendo que alguns modelos podem ser linearizáveis, principalmente utilizando-se a teoria de logaritmos. Aquelas funções que não podem ser linearizadas são ajustadas por regressão não linear. Ressalta-se que a teoria de regressão linear é mais consolidada que a regressão não linear.

Os modelos testados no presente estudo foram ajustadas por meio de regressão não linear, segundo o procedimento de Marquadt, pelo método de mínimos quadrados, com múltiplas iterações, de acordo com Marquadt (1963), utilizando-se o software *Table Curve 2d*.

A maioria dos algoritmos para estimar parâmetros não lineares pelo método de mínimos quadrados centraram-se sobre qualquer uma de duas aproximações. Por um lado, o modelo pode ser expandido como uma série de Taylor e as correções dos parâmetros calculadas a cada iteração na suposição de linearidades locais. Por outro lado, várias modificações do método gradiente estão sendo usadas. O método de Marquadt utiliza o método de máxima verossimilhança, que executa uma ótima interpolação entre o método da série de Taylor e o método gradiente. A interpolação é baseada em máxima verossimilhança, na qual a série de Taylor truncada

fornece uma representação adequada de um modelo não-linear (MARQUADT, 1963).

Seja $E(y) = f(x_1, x_2, \dots, x_m; \beta_1, \beta_2, \dots, \beta_k) = f(x, \beta)$ o modelo a ser ajustado aos dados, em que x_1, x_2, \dots, x_m são as variáveis independentes, $\beta_1, \beta_2, \dots, \beta_k$ são os k parâmetros e $E(y)$ o valor esperado da variável y . Denotando-se os pontos amostrados como $(Y_i, X_{1i}, X_{2i}, \dots, X_{mi})$, $i = 1, 2, \dots, n$. O problema consiste em computar essas estimativas dos parâmetros que minimizem $\phi = \sum_{i=1}^n [Y_i - \hat{Y}_i]^2 = \|Y - \hat{Y}\|^2$, em que \hat{Y}_i é o valor de y predito por $E(y)$ no i -ésimo ponto amostrado (MARQUADT, 1963).

MÉDIA E VARIÂNCIA DE UMA FUNÇÃO DENSIDADE DE PROBABILIDADE

A média e a variância podem ser determinadas por meio da Função Geratriz dos Momentos ou da Expectativa Matemática. Essa última foi utilizada no presente estudo.

EXPECTATIVA MATEMÁTICA

Seja X uma variável aleatória contínua com função densidade de probabilidade f . Defina-se o valor esperado $E(X)$ de X , em que $E(X)$ é a média, conforme Equação 17 a seguir (GUIDORIZZI, 1994):

$$E(X) = \int_{-\infty}^{+\infty} x \cdot f(x) dx \quad \text{EQUAÇÃO 17}$$

A variância pode ser escrita da seguinte forma (EQUAÇÃO 18):

$$\sigma_x^2 = \int_{-\infty}^{+\infty} x^2 \cdot f(x) dx - [E(X)]^2 \quad \text{EQUAÇÃO 18}$$

AValiação DO AJUSTE

Os modelos ajustados foram comparados por meio do coeficiente de determinação múltiplo, coeficiente de determinação múltiplo ajustado, erro padrão da estimativa recalculado em percentagem, teste do χ^2 , pelo teste de Kolmogorov-Smirnov e pela análise dos resíduos.

- COEFICIENTE DE DETERMINAÇÃO MÚLTIPLO

O coeficiente de determinação múltiplo é a razão entre a soma dos quadrados da regressão e a soma dos quadrados totais, de acordo com a Equação 19:

$$R^2 = \frac{SQ_{regressão}}{SQT} \quad \text{EQUAÇÃO 19}$$

- COEFICIENTE DE DETERMINAÇÃO MÚLTIPLO AJUSTADO

O Coeficiente de determinação múltiplo ajustado é a correção do R^2 para modelos com diferentes números de parâmetros (EQUAÇÃO 20).

$$R^2_{ajustado} = 1 - \left\{ (1 - R^2) \cdot \left(\frac{n-1}{n-p} \right) \right\} \quad \text{EQUAÇÃO 20}$$

- ERRO PADRÃO DA ESTIMATIVA RECALCULADO

O erro padrão da estimativa recalculado é a raiz quadrada do quociente entre: 1) o somatório do quadrado da diferença entre os valores observados e os estimados, e 2) a diferença entre o número de dados e o número de parâmetros estimados pelo modelo em questão, conforme Equação 21:

$$S_{yx \text{ recalculado}} = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n-p}} \quad \text{EQUAÇÃO 21}$$

O erro padrão da estimativa em percentagem é dado pela razão entre o erro padrão da estimativa e a média (EQUAÇÃO 22):

$$S_{yx} \% = \frac{S_{yx}}{\bar{x}} \cdot 100 \quad \text{EQUAÇÃO 22}$$

Em que:

y_i é frequência observada;

\hat{y}_i é a frequência estimada pelos modelos;

\bar{x} é a média aritmética;

n é o número de dados; e

p o número de parâmetros estimados pelo modelo em questão.

- KOLMOGOROV E SMIRNOV

O teste de aderência de Kolmogorov e Smirnov consiste em comparar a máxima diferença entre a frequência observada e a frequência estimada, dividida pelo número de observações, e pode ser obtido pela Equação 23:

$$d_{calc} = \frac{(\max |F_o(X) - F_e(X)|)}{n} \quad \text{EQUAÇÃO 23}$$

Em que $F_o(X)$ é a frequência observada acumulada, $F_e(X)$ é a frequência esperada acumulada, estimada pelo modelo e n o número de observações.

Se d_{calc} for menor que o valor limite tabelado, aceita-se o ajuste. Se n for menor que 50, para $\alpha = 5\%$, e $\alpha = 1\%$, determina-se o valor crítico por meio das Equações 24 e 25 a seguir.

$$\alpha = 5\% \rightarrow 1,36 \cdot \sqrt{\frac{n_1 + n_2}{n_1 \cdot n_2}} \quad \text{EQUAÇÃO 24}$$

$$\alpha = 1\% \rightarrow 1,63 \cdot \sqrt{\frac{n_1 + n_2}{n_1 \cdot n_2}} \quad \text{EQUAÇÃO 25}$$

- QUI-QUADRADO (χ^2)

O Qui-quadrado é um teste de aderência e é calculado pela Equação 27:

$$\chi^2 = \sum \left[\frac{(y_i - \hat{y}_i)^2}{\hat{y}_i} \right] \quad \text{EQUAÇÃO 27}$$

- ANÁLISE DE RESÍDUOS

Tem se tornado consenso que nenhuma equação deve ser escolhida para uso em processos preditivos antes de se performar uma análise gráfica de resíduos. Essa análise, embora um tanto subjetiva, é que realmente vai indicar se o ajuste foi bom ao longo de toda a estimativa, se não há tendenciosidade, se os resíduos são independentes e se a variância é homogênea em toda a extensão dos valores estimados.

Define-se resíduo em percentagem como a razão entre a diferença da frequência observada e a estimada pelo modelo ajustado e a frequência observada.

- DESVIO RELATIVO

Utilizou-se o desvio relativo entre as somas observada e estimada das frequências para comparar as funções probabilísticas ajustadas, de acordo com a equação 26.

$$DESVIO\% = \frac{f_{obs} - f_{est}}{f_{obs}} \cdot 100 \quad \text{EQUAÇÃO 26}$$

As frequências foram elevadas à primeira, segunda, terceira e quarta potências, adequando o que foi proposto por Maltamo et al (1995).

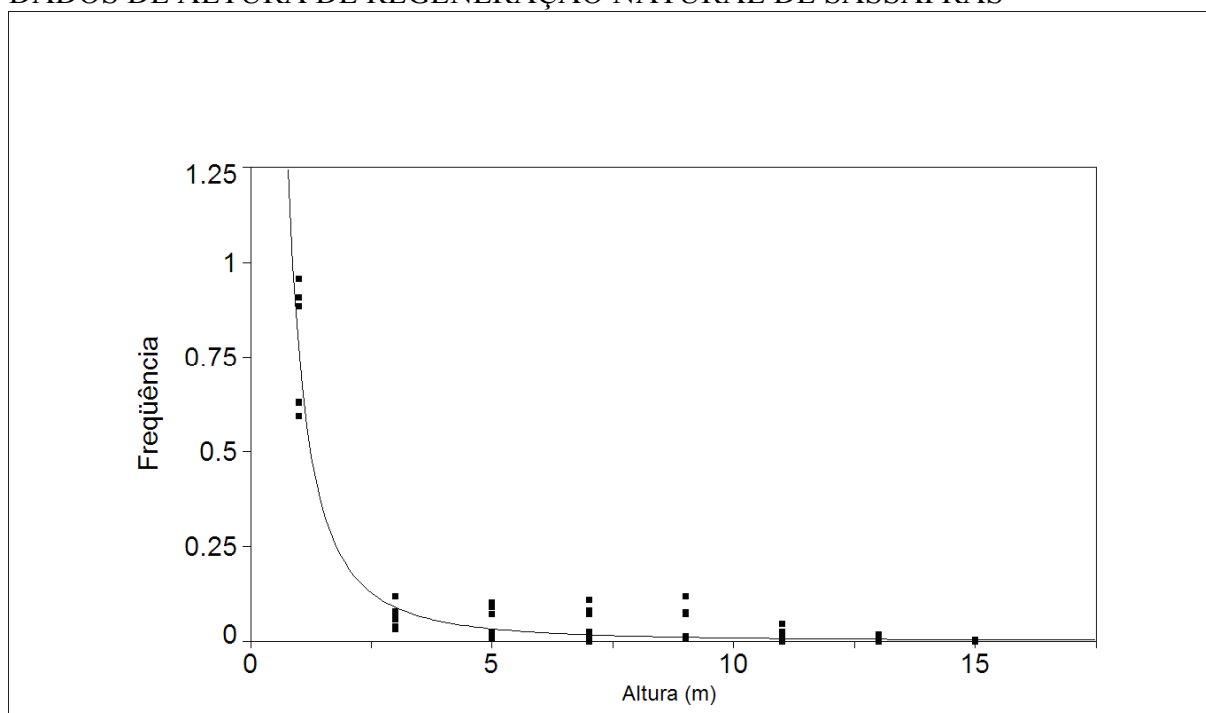
4 – RESULTADOS E DISCUSSÃO - DESENVOLVIMENTO DA FUNÇÃO DENSIDADE DE PROBABILIDADE PROPOSTA

Para desenvolver a função densidade de probabilidade proposta nesse trabalho, foram escolhidas 5 funções matemáticas, apresentadas a seguir.

No início escolheu-se uma equação que tivesse um comportamento não crescente, conforme ocorrente na distribuição exponencial, porém com maior flexibilidade (EQUAÇÃO 27).

$$f(x) = \frac{1}{(ax + b)^2} \quad \text{EQUAÇÃO 27}$$

FIGURA 14 – REPRESENTAÇÃO GRÁFICA DA EQUAÇÃO 27 AJUSTADO AOS DADOS DE ALTURA DE REGENERAÇÃO NATURAL DE SASSAFRÁS

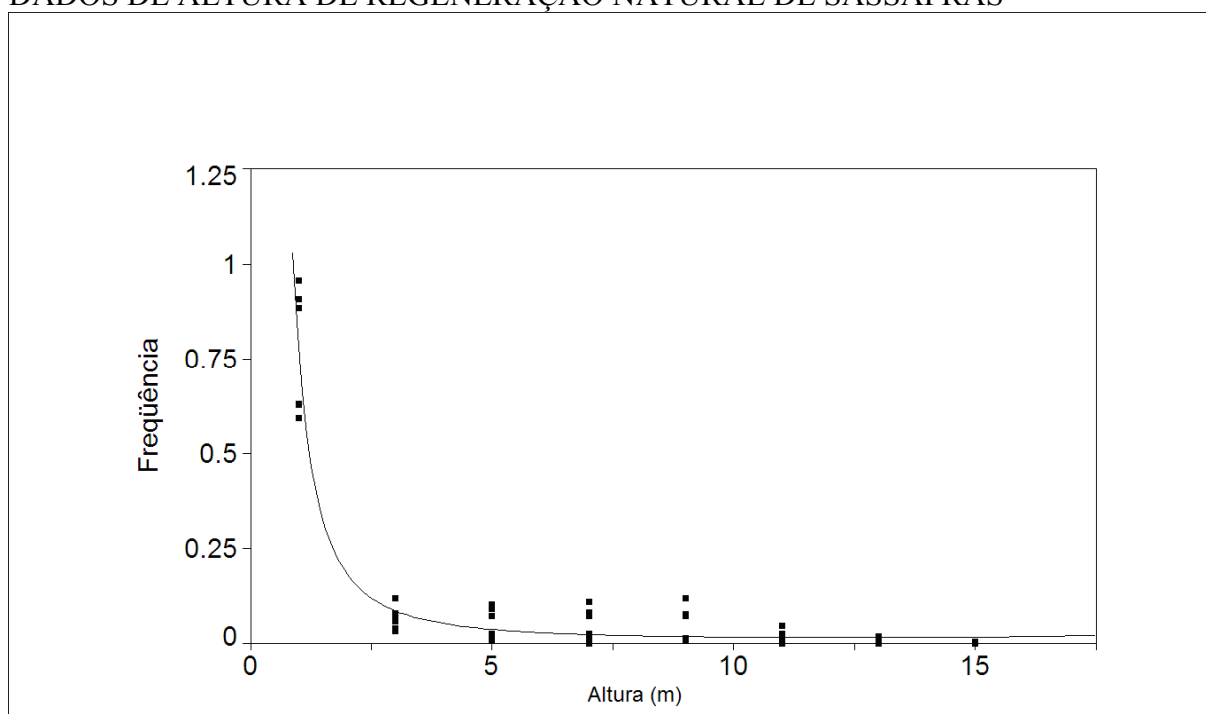


Nota-se na Figura 14 que o modelo ajustou-se de modo satisfatório, porém não atendeu a um dos pré-requisitos, que foi gerar uma equação unimodal.

Com o propósito de flexibilizar ainda mais a primeira equação, mais um coeficiente foi introduzido, resultando na Equação 28.

$$f(x) = \frac{1}{ax^2 + bx + c} \quad \text{EQUAÇÃO 28}$$

FIGURA 15 – REPRESENTAÇÃO GRÁFICA DA EQUAÇÃO 28 AJUSTADO AOS DADOS DE ALTURA DE REGENERAÇÃO NATURAL DE SASSAFRÁS



Este modelo gerou resultados semelhantes ao anterior e o mesmo defeito, a falta de uma moda (FIGURA 15).

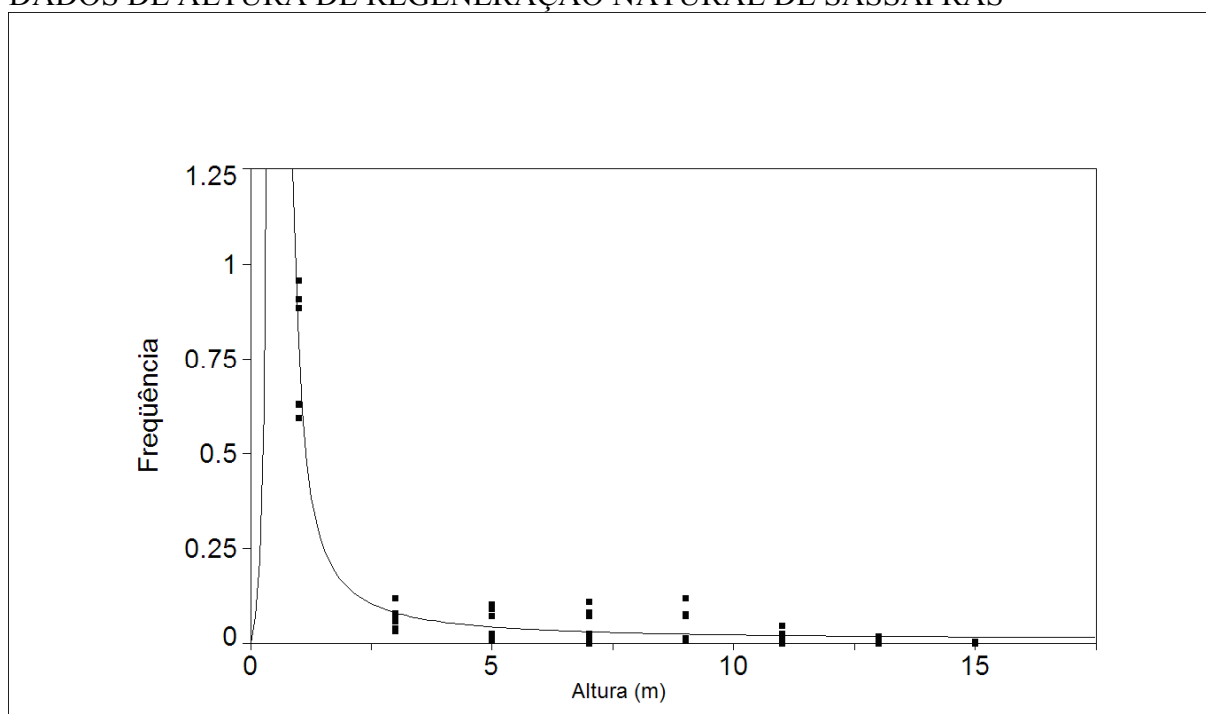
Nesta etapa do trabalho foi necessário rever as hipóteses e pesquisar novas famílias de curvas, mas sem alterar substancialmente a proposta inicial.

Com o intuito de arbitrar que a representação gráfica do modelo passasse pela origem (interseção dos eixos coordenados), a Equação 28 foi multiplicada por x (EQUAÇÃO 29). Dessa forma, quando $x = 0$, $f(x) = 0$. Para que não houvesse indeterminação, arbitrou-se que o coeficiente c fosse maior ou igual a zero.

$$f(x) = \frac{x}{ax^2 + bx + c} \quad \text{EQUAÇÃO 29}$$

Para verificar o comportamento da curva o termo multiplicador x e cada membro do denominador, exceto o termo independente, foram elevados ao quadrado, resultando na Equação 30. Esse procedimento foi adotado com a finalidade de tornar a distribuição mais “comportada”, ou seja, com amplitude menor, porém, o resultado foi pior que a do modelo testado anteriormente, resultando em um gráfico decrescente.

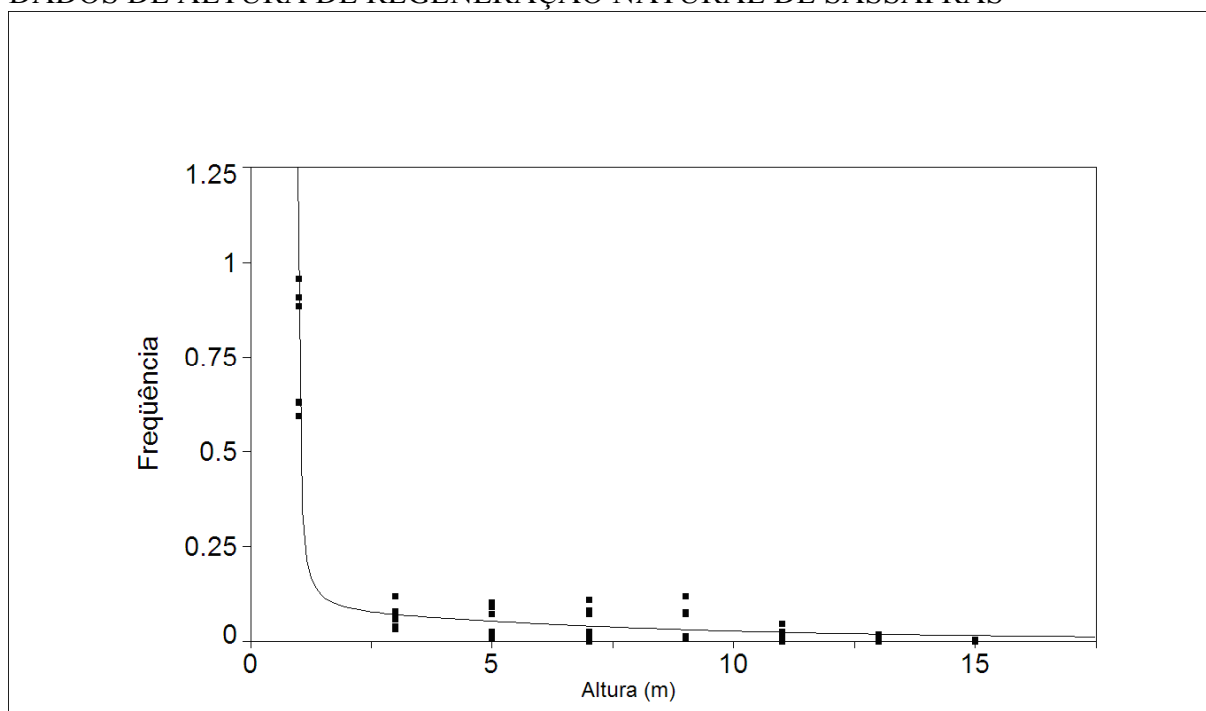
FIGURA 16 - REPRESENTAÇÃO GRÁFICA DA EQUAÇÃO 29 AJUSTADO AOS DADOS DE ALTURA DE REGENERAÇÃO NATURAL DE SASSAFRÁS



Observa-se na Figura 16 que a Equação 29 ajustou-se aos dados, gerando uma curva unimodal, porém gera valores para $f(x)$ altos, quando x aproxima-se da moda.

$$f(x) = \frac{x^2}{ax^4 + bx^2 + c} \quad \text{EQUAÇÃO 30}$$

FIGURA 17 - REPRESENTAÇÃO GRÁFICA DA EQUAÇÃO 30 AJUSTADO AOS DADOS DE ALTURA DE REGENERAÇÃO NATURAL DE SASSAFRÁS

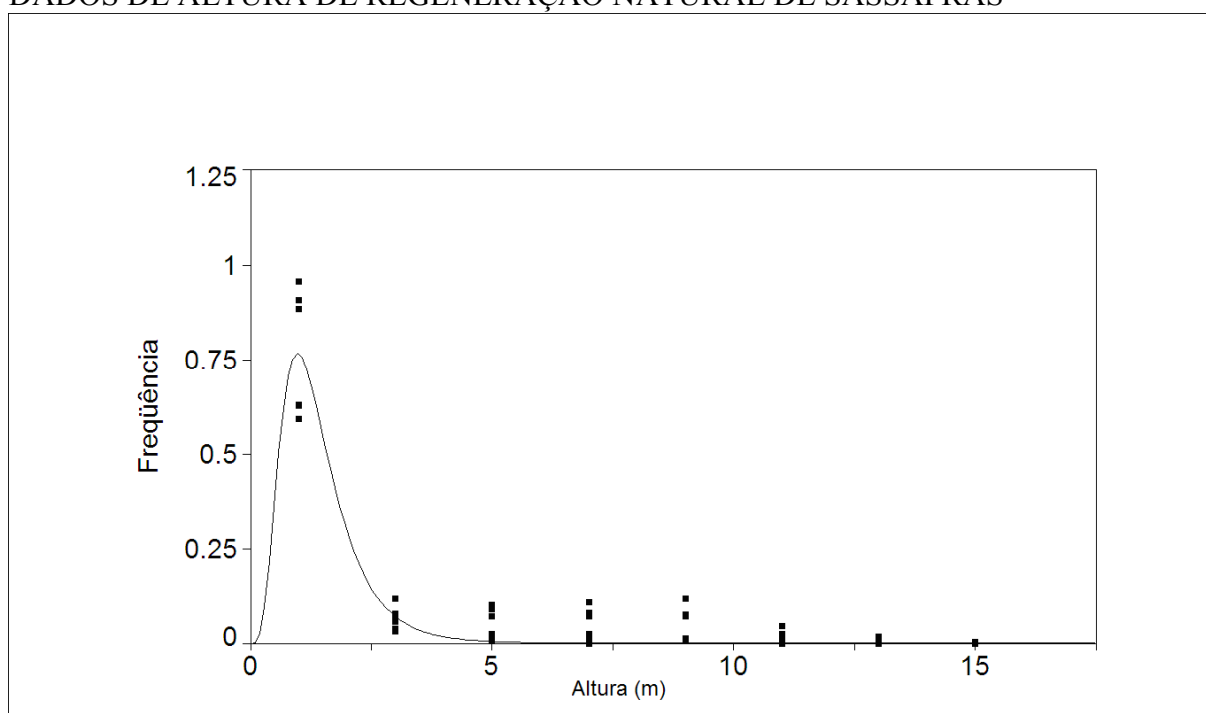


Na Figura 17 observa-se que a equação resulta no mesmo problema das duas primeiras, ausência de moda.

Com a inserção de parênteses no denominador e a flexibilização dos coeficientes desse e do numerador foi possível obter um modelo com todas as características desejadas, como pode ser observado na Figura 17.

$$f(x) = \frac{x^a}{(b + c \cdot x)^d} \quad \text{EQUAÇÃO 31}$$

FIGURA 18 - REPRESENTAÇÃO GRÁFICA DA EQUAÇÃO 31 AJUSTADO AOS DADOS DE ALTURA DE REGENERAÇÃO NATURAL DE SASSAFRÁS



Essa equação atendeu a todas as exigências propostas e foi, conseqüentemente, a escolhida (FIGURA 18).

A função $f(x)$, cujo domínio é o conjunto dos números reais positivos (R^+), ou seja, $0 \leq x < \infty$, $d \geq a+1$; $d \neq a+2$, $d \neq a+3$, $a > 0$, $b > 0$ e $c > 0$, definida a seguir, será considerada uma Função Densidade Probabilidade (*fdp*) se contemplar as condições 1, 2, 3 e 4:

$$f(x) = \frac{x^a}{(b + c \cdot x)^d}$$

CONDIÇÕES PARA QUE UMA FUNÇÃO SEJA CONSIDERADA UMA FDP:

1. A função deve possuir um domínio real e positivo. Assim, quando $x = 0$, $f(0) = 0$

$$f(x) = \frac{x^a}{(b + c \cdot x)^d}$$

$$f(0) = \frac{0^a}{(b + c \cdot 0)^d}$$

$$f(0) = 0$$

2. A função deve ser assintótica no eixo x , ou seja, $\lim_{x \rightarrow \infty} f(x) = 0$

$$\lim_{x \rightarrow \infty} \frac{x^a}{(b + c \cdot x)^d} = 0 \text{ pois}$$

Como o valor de d é sempre maior que $a + 2$, por definição, o denominador tenderá a infinito mais rapidamente que o numerador, logo, quando x tende a infinito, a função tende a 0.

3. A função deve ser integrável, ou seja, a integral da função, no intervalo da variável x entre 0 (zero) e infinito, deve tender para um valor real.

$$\int_0^{\infty} f(x) \cdot dx = \int_0^{\infty} \frac{x^a}{(b + c \cdot x)^d} \cdot dx$$

$$\int_0^{\infty} f(x) \cdot dx = \frac{\left(\frac{b}{c}\right)^a \cdot b^{(1-d)} \cdot \Gamma(d-1-a) \cdot \Gamma(a+1)}{c \cdot \Gamma(d)}$$

Logo, a *fdp* será o produto entre a função original e o inverso do resultado anterior:

$$f_{dp}(x) = \frac{x^a \cdot c \cdot \Gamma(d)}{(b + c \cdot x)^d \cdot \left(\frac{b}{c}\right)^a \cdot b^{(1-d)} \cdot \Gamma(d-1-a) \cdot \Gamma(a+1)} \quad \text{EQUAÇÃO 32}$$

A partir da *fdp* apresentada anteriormente surgem as restrições de a e d , $d > a + 1$ e $d > 0$.

4. A integral da *fdp*, no intervalo da variável x entre 0 (zero) e infinito, deve ser igual a 1 (um).

$$\int_0^{\infty} f_{dp}(x) \cdot dx = \int_0^{\infty} \frac{x^a \cdot c \cdot \Gamma(d)}{(b + c \cdot x)^d \cdot \left(\frac{b}{c}\right)^a \cdot b^{(1-d)} \cdot \Gamma(d-1-a) \cdot \Gamma(a+1)} \cdot dx$$

$$\int_0^{\infty} fdp(x) \cdot dx = 1$$

MÉDIA ARITMÉTICA

A média aritmética de uma função probabilística pode ser obtida por meio da expectativa matemática, ou seja, a média é o resultado da integral do produto da função probabilística por x , no intervalo da variável x entre 0 (zero) e infinito.

$$\int_0^{\infty} x \cdot fdp(x) \cdot dx = \int_0^{\infty} x \cdot \frac{x^a \cdot c \cdot \Gamma(d)}{(b + c \cdot x)^d \cdot \left(\frac{b}{c}\right)^a \cdot b^{(1-d)} \cdot \Gamma(d-1-a) \cdot \Gamma(a+1)} \cdot dx$$

$$\int_0^{\infty} x \cdot fdp(x) \cdot dx = \frac{b(a+1)}{c(d-2-a)}$$

$$\text{Média aritmética} = \frac{b(a+1)}{c(d-2-a)}$$

EQUAÇÃO 33

VARIÂNCIA

A variância, da mesma maneira que a média aritmética pode ser obtida por meio da expectativa matemática, ou seja, é a diferença entre o resultado da integral do produto da função probabilística por x^2 e o quadrado da média aritmética.

$$\int_0^{\infty} x^2 \cdot fdp(x) \cdot dx = \int_0^{\infty} x^2 \cdot \frac{x^a \cdot c \cdot \Gamma(d)}{(b + c \cdot x)^d \cdot \left(\frac{b}{c}\right)^a \cdot b^{(1-d)} \cdot \Gamma(d-1-a) \cdot \Gamma(a+1)} \cdot dx$$

$$\int_0^{\infty} x^2 \cdot fdp(x) \cdot dx = \frac{b^2(a+1)(a+2)}{c^2(d-3-a)(d-2-a)}$$

$$\int_0^{\infty} x^2 \cdot fdp(x) \cdot dx - \left(\int_0^{\infty} x \cdot fdp(x) \cdot dx \right)^2 = \frac{b^2(a+1)(a+2)}{c^2(d-3-a)(d-2-a)} - \left(\frac{b(a+1)}{c(d-2-a)} \right)^2$$

$$\text{Variância} = - \frac{(a+1)b^2(d-1)}{c^2(a+3-d)(a+2-d)^2}$$

EQUAÇÃO 34

MODA

A moda representa o ponto, ou os pontos, de maior frequência, em uma fdp , e pode ser obtido a partir da primeira derivada da função em relação à x e igualando-se o resultado a 0

(zero), ou seja, $\frac{df(x)}{dx} = 0$.

$$\frac{df(x)}{dx} = \frac{ax^a}{x(b+cx)^d} - \frac{cdx^a}{(b+cx)^{d+1}}$$

$$\frac{df(x)}{dx} = 0$$

$$0 = \frac{ax^a}{x(b+cx)^d} - \frac{cdx^a}{(b+cx)^{d+1}}$$

$$x = \frac{ab}{c(d-a)}, \text{ no caso de a função ser unimodal.}$$

EQUAÇÃO 35

PONTOS DE INFLEXÃO

Os pontos de inflexão, se existirem, podem ser obtidos calculando-se o valor de x ao igualar a zero a segunda derivada de uma função, ou seja, $\frac{d^2f(x)}{dx^2} = 0$.

$$\frac{d^2f(x)}{dx^2} = \frac{a^2x^2}{x^2(b+cx)^d} - \frac{ax^2}{x^2(b+cx)^d} - \frac{2acdx^a}{x(b+cx)^{d+1}} + \frac{c^2d^2x^a}{(b+cx)^{d+2}} + \frac{c^2dx^a}{(b+cx)^{d+2}}$$

$$\frac{d^2f(x)}{dx^2} = 0$$

$$0 = \frac{a^2x^2}{x^2(b+cx)^d} - \frac{ax^2}{x^2(b+cx)^d} - \frac{2acdx^a}{x(b+cx)^{d+1}} + \frac{c^2d^2x^a}{(b+cx)^{d+2}} + \frac{c^2dx^a}{(b+cx)^{d+2}}$$

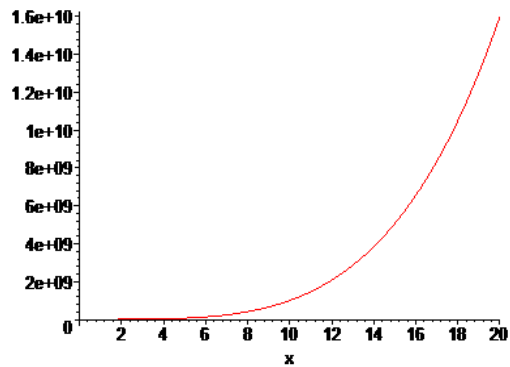
$$x = \frac{b}{c} \cdot \frac{a^2 - ad - a \pm \sqrt{-ad(a-d-1)}}{(a-d-(a-d)^2)}$$

EQUAÇÃO 36

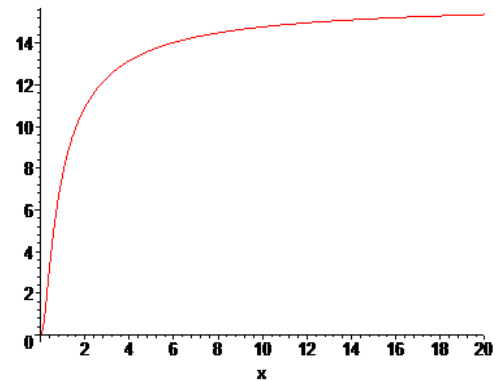
A referida função não terá pontos de inflexão quando: I) $c = 0$, II) $a - d - (a - d)^2 = 0$ ou III) $-ad(a-d-1) < 0$. Quando $-ad(a-d-1) = 0$ haverá apenas um ponto de inflexão. No caso de a função ser unimodal, haverá dois pontos de inflexão, e isso ocorrerá quando $-ad(a-d-1) > 0$.

Os resultados obtidos por meio da avaliação das condições de existência de pontos de inflexão podem ser observados na Figura 19.

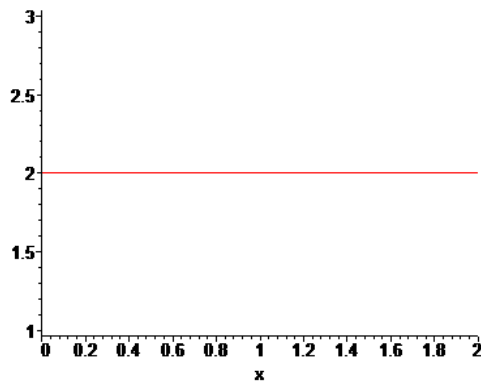
FIGURA 19 – REPRESENTAÇÃO GRÁFICA DO MODELO PROPOSTO, VARIANDO-SE OS COEFICIENTES, DE ACORDO COM AS CONCLUSÕES DO ESTUDO SOBRE OS PONTOS DE INFLEXÃO



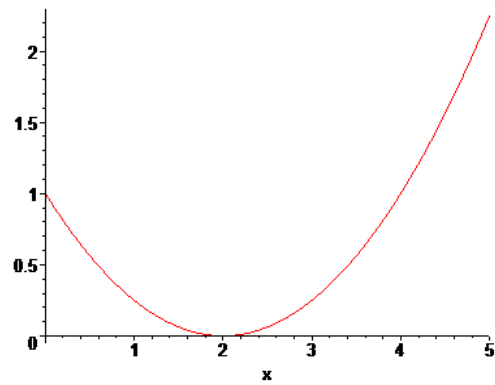
Sem pontos de inflexão ($c = 0$)



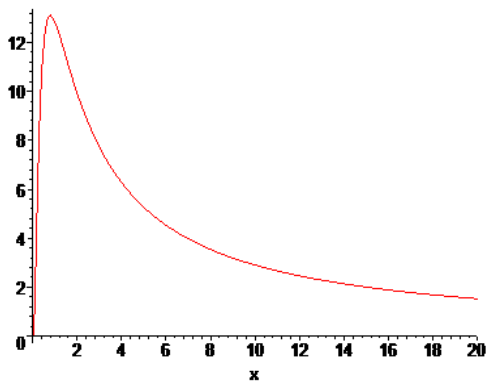
Sem pontos de inflexão
($a - d - (a - d)^2 = 0$)



Sem pontos de inflexão ($-ad(a - d - 1) < 0$)



Um ponto de inflexão ($-ad(a - d - 1) = 0$)



Dois pontos de inflexão $-ad(a - d - 1) > 0$

Para que a *fdp* exista é necessário que as seguintes relações se verifiquem:

1. $b > 0$ e $c > 0$. Essa condição é imposta pela própria função, $f(x) = \frac{x^a}{(b + c \cdot x)^d}$, pois, se b e c forem iguais a 0 (zero), o denominador torna-se 0 (zero) e a função tenderá para o infinito, o que não é desejável;
2. $d \geq 0$, $d \geq a + 1$ e $a \geq -1$. Como a função gamma é definida apenas no campo dos números reais positivos, os coeficientes devem seguir as restrições anteriores, ao considerar a *fdp*,

$$f_{dp}(x) = \frac{x^a \cdot c \cdot \Gamma(d)}{(b + c \cdot x)^d \cdot \left(\frac{b}{c}\right)^a \cdot b^{(1-d)} \cdot \Gamma(d-1-a) \cdot \Gamma(a+1)} ;$$

3. $c > 0$ e $d \neq a + 2$. Caso a restrição anterior não seja verificada a média, $\frac{b(a+1)}{c(d-2-a)}$, tenderá para o infinito;
4. $c > 0$, $d \neq a + 2$ e $d \neq a + 3$. Se essas condições não forem consideradas, o denominador da expressão, $\frac{(a+1)b^2(d-1)}{c^2(a+3-d)(a+2-d)^2}$, será igual a zero e a variância tenderá para infinito;
5. $c > 0$ e $d \neq a$. Para que haja moda, $x = \frac{ab}{c(d-a)}$, essas exigências devem ser atendidas; e
6. $a > 0$, $b > 0$ e $c > 0$. Essas três últimas restrições tornam a função probabilística no domínio determinado.

Assim, combinando as restrições anteriores tem-se:

- $d \geq a + 1$, $d \neq a + 2$ e $d \neq a + 3$;
- a, b e $c > 0$.

TENDÊNCIA DO MODELO DESENVOLVIDO A NORMALIDADE

Ao considerar-se que a média é igual à moda, assume-se que a função tende à normalidade.

➤ Média $x = \frac{b(a+1)}{c(d-2-a)}$

- $\text{Moda } x = \frac{ab}{c(d-a)}$
- $\frac{b(a+1)}{c(d-2-a)} = \frac{ab}{c(d-a)}$
- $\frac{ab+b}{cd-2c-ac} = \frac{ab}{cd-ac}$
- $(ab+b)(cd-ac) = ab(cd-2c-ac)$
- $abcd - a^2bc + bcd - abc = abcd - 2abc - a^2bc$
- $bcd + abc = 0$
- $bc(d+a) = 0$

Logo,

- $d = -a$. Como $d \geq a+1$, essa hipótese não ocorrerá.
- $b = 0$. Como $b > 0$, essa hipótese não ocorrerá.
- $c = 0$. Como $c > 0$, essa hipótese não ocorrerá.

FORMAS QUE A FUNÇÃO PODE ASSUMIR

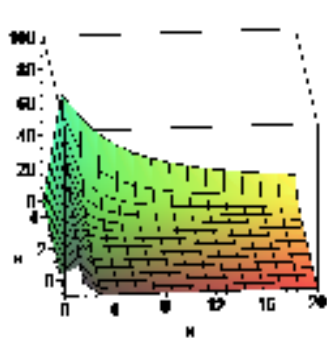
A forma e amplitude da representação gráfica da função variam de acordo com a combinação dos coeficientes.

As formas que o modelo proposto pode assumir estão representadas graficamente na Figura 20, de forma tridimensional.

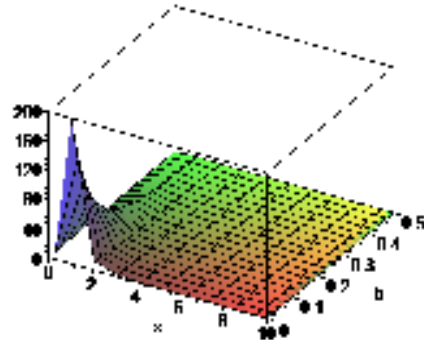
Como se sabe, a forma de um modelo se dá de acordo com seus coeficientes e, ao considerá-los como sendo iguais, pode-se observar a forma que o modelo aqui proposto assumirá (FIGURA 20).

Para visualizar-se mais claramente as formas que o modelo proposto pode assumir, foram representados os gráficos, fixando-se três coeficientes e variando outro (FIGURA 21; 22; 23; 24; 25).

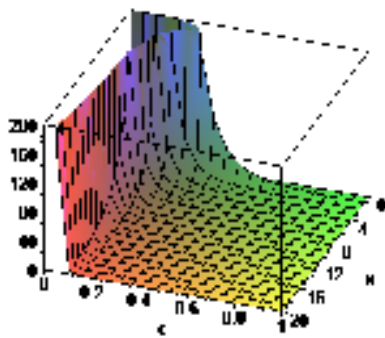
FIGURA 20 – REPRESENTAÇÃO TRIDIMENSIONAL DAS FORMAS GRÁFICAS ASSUMIDAS PELO MODELO PROPOSTO



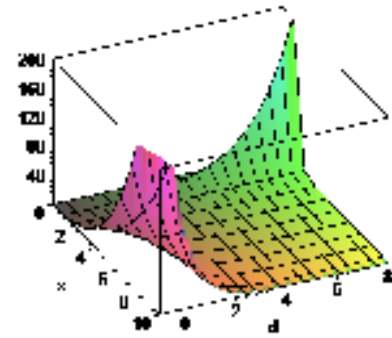
a variando entre -1 e 5; $b = 0,1$; $c = 0,4$; e $d = 6$



b variando entre 0 e 0,5; $a = 2$; $c = 0,4$ e $d = 6$



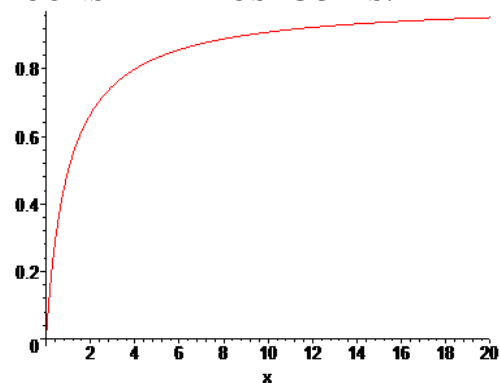
c variando entre 0 e 1; $a = 2$; $b = 0,1$; e $d = 6$



d variando entre -1 e 5; $a = 2$; $b = 0,1$; e $c = 0,4$

FIGURA 21 – REPRESENTAÇÃO GRÁFICA DO MODELO PROPOSTO, EM QUE OS COEFICIENTES SÃO ARBITRARIAMENTE CONSIDERADOS IGUAIS.

Indeterminação, divisão por zero.

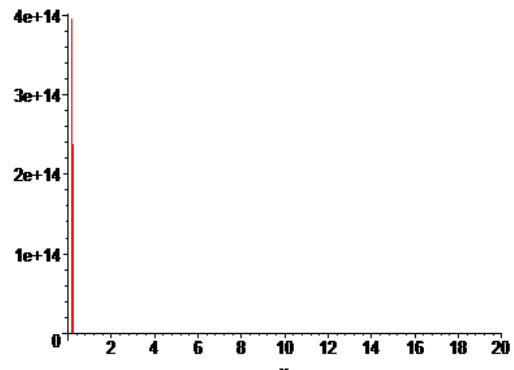


$$a = b = c = d = 0 \quad f(x) = \frac{1}{0}$$

$$a = b = c = d = 1 \quad f(x) = \frac{x}{(1+x)} \text{ Hipérbole}$$

FIGURA 22 – REPRESENTAÇÃO GRÁFICA DO MODELO PROPOSTO, COM a , c E d FIXOS E VARIANDO b

Indeterminação, divisão por zero.

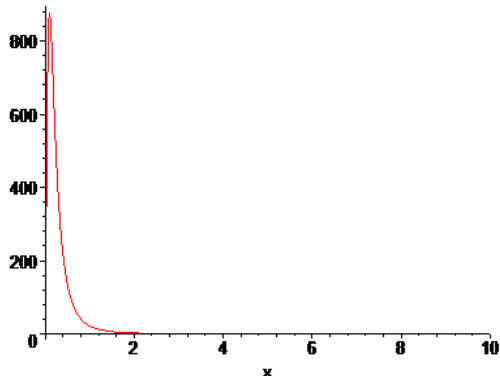


$$a = 2, b = 0, c = 0, d = 6$$

$$f(x) = \frac{x^2}{0}$$

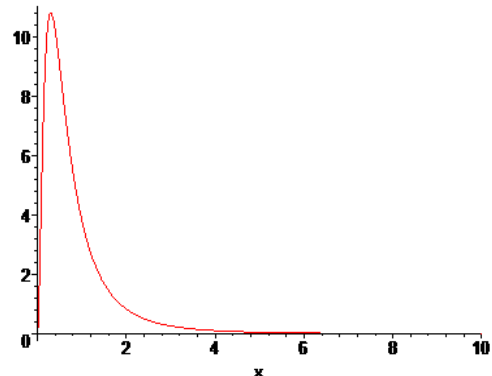
$$a = 2, b = -0,1, c = 0,5, d = 6$$

$$f(x) = \frac{x^2}{(-0,1 + 0,5 \cdot x)^6} \text{ Unimodal}$$



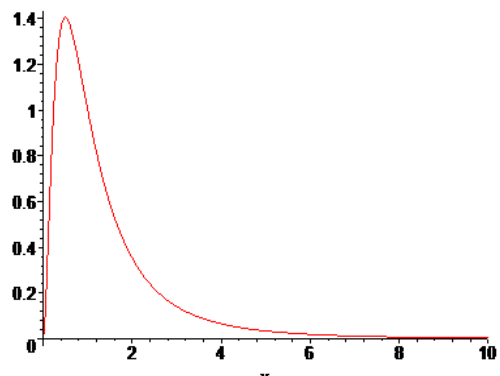
$$a = 2, b = 0,1, c = 0,5, d = 6$$

$$f(x) = \frac{x^2}{(0,1 + 0,5 \cdot x)^6} \text{ Unimodal}$$



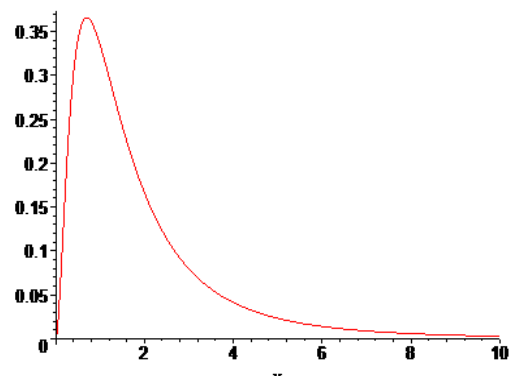
$$a = 2, b = 0,3, c = 0,5, d = 6$$

$$f(x) = \frac{x^2}{(0,3 + 0,5 \cdot x)^6} \text{ Unimodal}$$



$$a = 2, b = 0,5, c = 0,5, d = 6$$

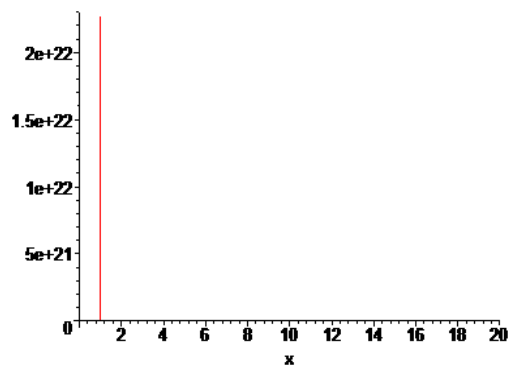
$$f(x) = \frac{x^2}{(0,5 + 0,5 \cdot x)^6} \text{ Unimodal}$$



$$a = 2, b = 0,7, c = 0,5, d = 6$$

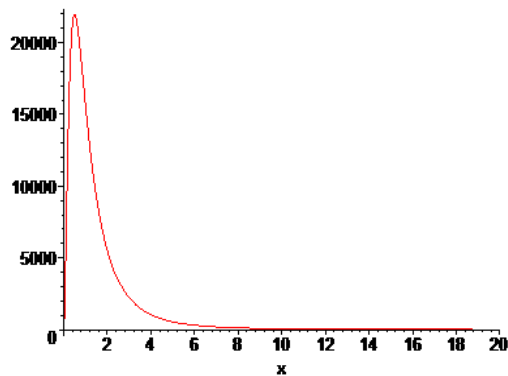
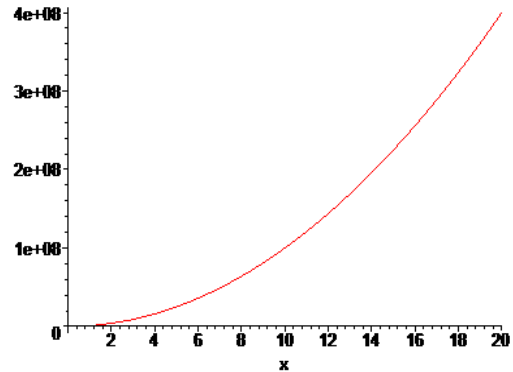
$$f(x) = \frac{x^2}{(0,7 + 0,5 \cdot x)^6} \text{ Unimodal}$$

FIGURA 23 – REPRESENTAÇÃO GRÁFICA DO MODELO PROPOSTO, COM a , b E d FIXOS E VARIANDO c



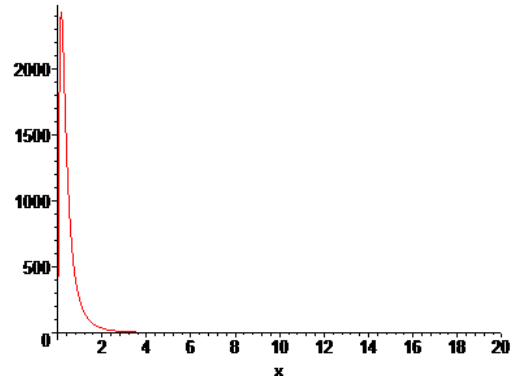
$$a = 2, b = 0,1, c = -0,1, d = 6 \quad a = 2, b = 0,1, c = 0, d = 6$$

$f(x) = \frac{x^2}{(0,1 - 0,1 \cdot x)^6}$, quando $x = 1$ haverá uma indeterminação. $f(x) = \frac{x^2}{(0,1)^6}$ Função do segundo grau.



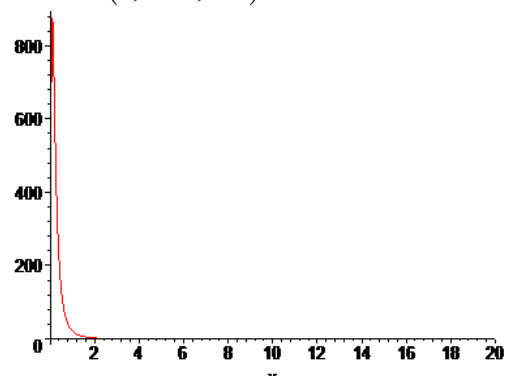
$$a = 2, b = 0,1, c = 0,1, d = 6$$

$f(x) = \frac{x^2}{(0,1 + 0,1 \cdot x)^6}$ Unimodal



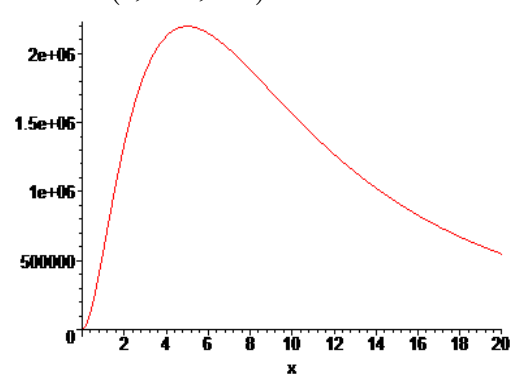
$$a = 2, b = 0,1, c = 0,3, d = 6$$

$f(x) = \frac{x^2}{(0,1 + 0,3 \cdot x)^6}$ Unimodal



$$a = 2, b = 0,1, c = 0,5, d = 6$$

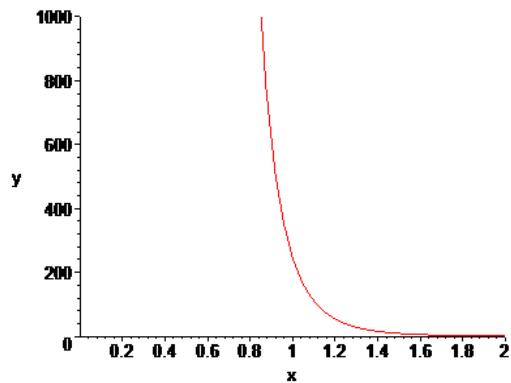
$f(x) = \frac{x^2}{(0,1 + 0,5 \cdot x)^6}$ Unimodal



$$a = 2, b = 0,1, c = 0,01, d = 6$$

$f(x) = \frac{x^2}{(0,1 + 0,01 \cdot x)^6}$ Unimodal

FIGURA 24 – REPRESENTAÇÃO GRÁFICA DO MODELO PROPOSTO, COM b, c E d FIXOS E VARIANDO a

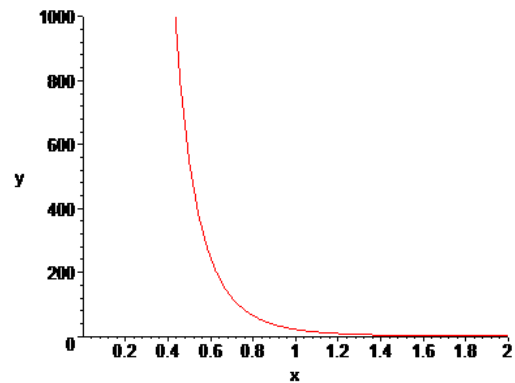


$$a = -1, b = 0,1, c = 0,5, d = 6$$

$$f(x) = \frac{x^{-1}}{(0,1 + 0,5 \cdot x)^6}$$

negativa.

Exponencial

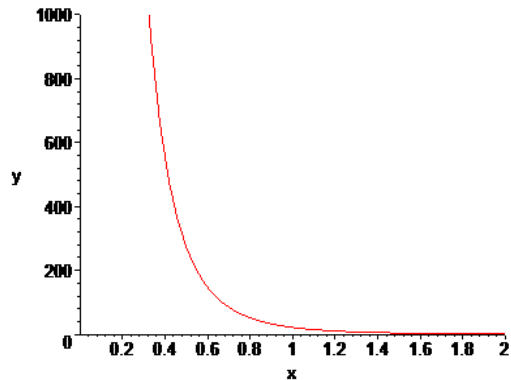


$$a = 0, b = 0,1, c = 0,5, d = 6$$

$$f(x) = \frac{1}{(0,1 + 0,5 \cdot x)^6}$$

negativa.

Exponencial

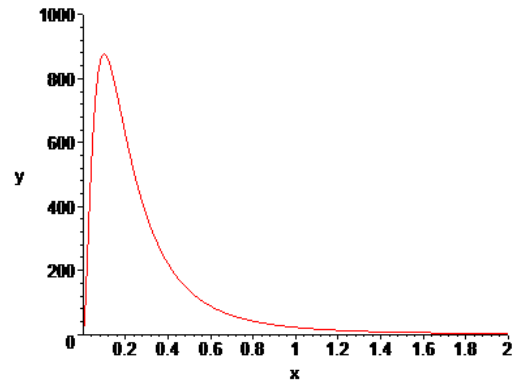


$$a = 1, b = 0,1, c = 0,5, d = 6$$

$$f(x) = \frac{x}{(0,1 + 0,5 \cdot x)^6}$$

negativa.

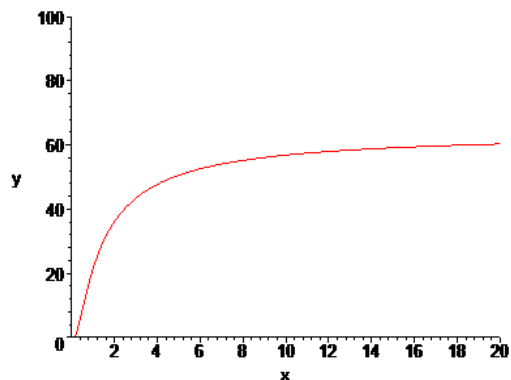
Exponencial



$$a = 2, b = 0,1, c = 0,5, d = 6$$

$$f(x) = \frac{x^2}{(0,1 + 0,5 \cdot x)^6}$$

Unimodal

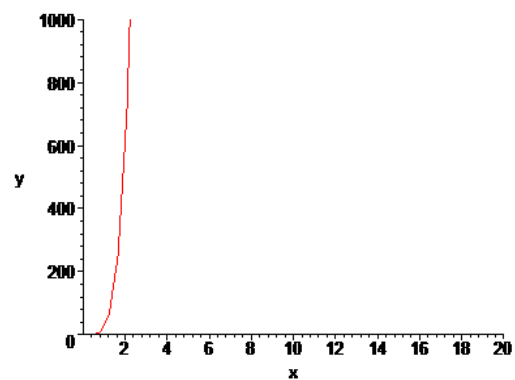


$$a = 6, b = 0,1, c = 0,5, d = 6$$

$$f(x) = \frac{x^6}{(0,1 + 0,5 \cdot x)^6}$$

em um valor de y

Função assintótica



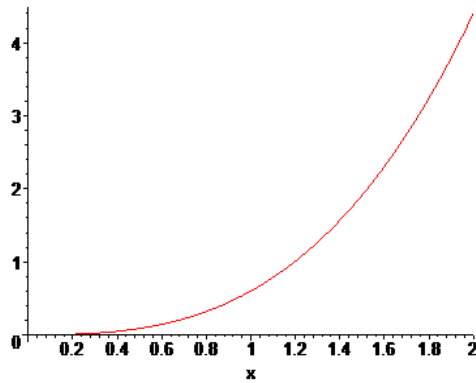
$$a = 10, b = 0,1, c = 0,5, d = 6$$

$$f(x) = \frac{x^{10}}{(0,1 + 0,5 \cdot x)^6}$$

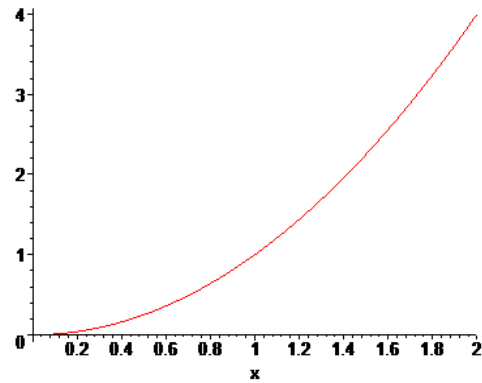
positiva

Exponencial

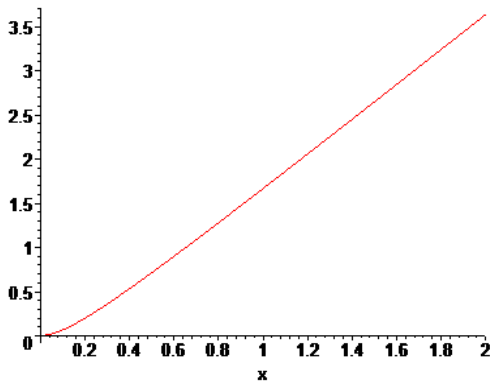
FIGURA 25 – REPRESENTAÇÃO GRÁFICA DO MODELO PROPOSTO, COM a , b E c FIXOS E VARIANDO d



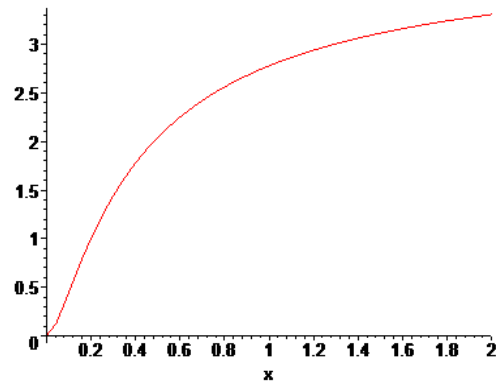
$a = 2, b = 0,1, c = 0,5, d = -1$
 $f(x) = \frac{x^2}{(0,1 + 0,5 \cdot x)^{-1}}$ Função crescente



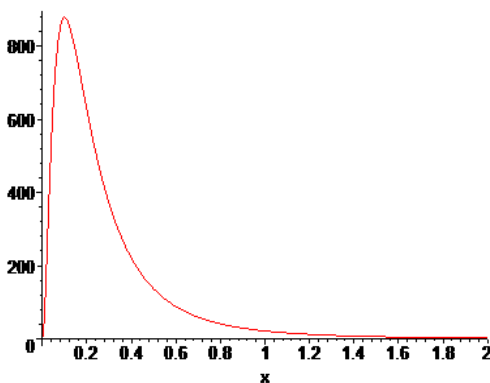
$a = 2, b = 0,1, c = 0,5, d = 0$
 $f(x) = x^2$ Função do segundo grau.



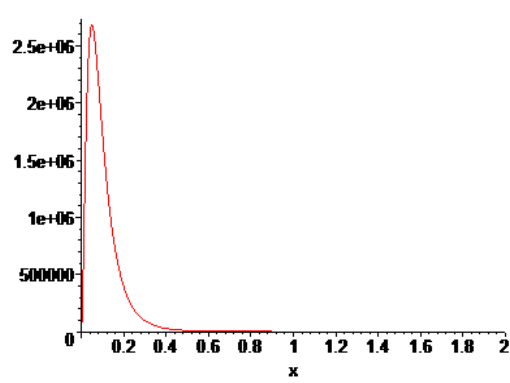
$a = 2, b = 0,1, c = 0,5, d = 1$
 $f(x) = \frac{x^2}{(0,1 + 0,5 \cdot x)^1}$ Função crescente



$a = 2, b = 0,1, c = 0,5, d = 2$
 $f(x) = \frac{x^2}{(0,1 + 0,5 \cdot x)^2}$ Função crescente



$a = 2, b = 0,1, c = 0,5, d = 6$
 $f(x) = \frac{x^2}{(0,1 + 0,5 \cdot x)^6}$ Unimodal



$a = 2, b = 0,1, c = 0,5, d = 10$
 $f(x) = \frac{x^2}{(0,1 + 0,5 \cdot x)^{10}}$ Unimodal

Assim, conclui-se que:

- a) b desloca a função ao longo do eixo x ;
- b) c altera a forma da função, variando sua dispersão ao longo do eixo x , e acumula toda a probabilidade em um único ponto, quando assume valor 0 (zero), se b for diferente de 0 zero);
- c) a e d alteram a forma
- d) Todos os coeficientes contribuem para a média, a variância, a moda e para os pontos de inflexão.

5 - SÍNTESE DO MODELO PROPOSTO

O modelo proposto nesse trabalho é uma função $f(x)$, cujo domínio é o conjunto dos números reais positivos (R^+), ou seja, $0 \leq x < \infty$, $d \geq a+1$, $d \neq a+2$, $d \neq a+3$, $a > 0$, $b > 0$ e $c > 0$, definido a seguir:

$$f(x) = \frac{x^a}{(b + c \cdot x)^d}$$

A média aritmética e variância podem ser calculadas pelas expressões:

$$\text{Média aritmética} = \frac{b(a+1)}{c(d-2-a)}$$

$$\text{Variância} = -\frac{(a+1)b^2(d-1)}{c^2(a+3-d)(a+2-d)^2}$$

O desvio padrão é obtido extraíndo-se a raiz quadrada da variância, como se segue:

$$\text{Desvio padrão} = \frac{b}{c(a+2-d)} \sqrt{-\frac{(a+1)(d-1)}{(a+3-d)}}$$

A moda e os pontos de inflexão, quando existirem, poderão ser calculados utilizando-se as seguintes fórmulas:

$$\text{Moda} = \frac{ab}{c(d-a)}$$

$$\text{Pontos de inflexão} = \frac{b}{c} \cdot \frac{a^2 - ad - a \pm \sqrt{-ad(a-d-1)}}{(a-d-(a-d)^2)}$$

6 - APLICAÇÃO/VERIFICAÇÃO

Para a verificação da eficiência do modelo proposto neste trabalho, esse e as principais distribuições utilizadas no meio florestal foram ajustadas e comparadas entre si. Foi utilizada a distribuição de dados da altura da regeneração natural de sassafrás das seis (6) parcelas permanentes do PELD, medidas durante o Projeto Sassafrás, as quais foram agrupadas em classes de dois metros. Para efetuar os testes de aderência os dados agrupados foram transformados, dividindo-se cada frequência pelo total de indivíduos de cada parcela. Os resultados estão contidos na Tabela 01.

TABELA 01 – DISTRIBUIÇÃO DE FREQUÊNCIAS DA ALTURA DA REGENERAÇÃO NATURAL DE SASSAFRÁS

Altura Total	X_i	Frequência Acumulada	Frequência Absoluta	Frequência Relativa
00 - 02	1	1468	1468	0,7583
02 - 04	3	1594	126	0,0651
04 - 06	5	1695	101	0,0522
06 - 08	7	1793	98	0,0506
08 - 10	9	1892	99	0,0511
10 - 12	11	1926	34	0,0176
12 - 14	13	1934	8	0,0041
14 - 16	15	1936	2	0,0010

Os coeficientes das principais distribuições probabilísticas, bem como da função desenvolvida no presente trabalho, estão apresentados na Tabela 02.

TABELA 02 – COEFICIENTES AJUSTADOS

Coeficientes	Normal	Weibull	Gamma	Beta	Exponencial	Proposto
a	2,2913	0,0000	0,7419	0,4925	3,4953	2,0000
b	2,6601	1,7706	3,0883	2,9464		0,2000
c		0,4218		0,0000		0,1161
d				16,0000		6,2550

Os coeficientes apresentados na tabela 02 foram substituídos nas respectivas funções (representadas a seguir) e os respectivos gráficos estão contidos na Figura 26.

NORMAL

$$f(x) = \begin{cases} \frac{1}{\sqrt{2\pi} \cdot 2,6601} \cdot e^{-\frac{1}{2} \left(\frac{x-2,2913}{2,6601} \right)^2} & -\infty < x < +\infty \\ 0 & \text{e.o.c.} \end{cases} \quad \text{EQUAÇÃO 37}$$

WEIBULL

$$f(x) = \begin{cases} \frac{0,4218}{7,1388} \cdot \left(\frac{x}{7,1388} \right)^{0,4218-1} \cdot e^{-\left(\frac{x}{7,1388} \right)^{0,4218}} & \text{se } x \geq a, b > 0 \text{ e } c > 0 \\ 0 & \text{e.o.c.} \end{cases} \quad \text{EQUAÇÃO 38}$$

GAMMA

$$f(x) = \begin{cases} \frac{x^{0,7419-1} \cdot e^{-x/3,0883}}{3,0883^{0,7419} \cdot \Gamma(0,7419)} & \text{para } x \geq 0 \\ 0 & \text{e.o.c.,} \end{cases} \quad \text{EQUAÇÃO 39}$$

BETA

$$f(x) = \begin{cases} \frac{\Gamma(0,4925 + 2,9464)}{\Gamma(0,4925) \cdot \Gamma(2,9464) \cdot (16-0)^{0,4925+2,9464-1}} \cdot (x-0)^{0,4925-1} \cdot (16-x)^{2,9464-1} & \text{se } a < x < b \\ 0 & \text{e.o.c.} \end{cases}$$

EQUAÇÃO 36

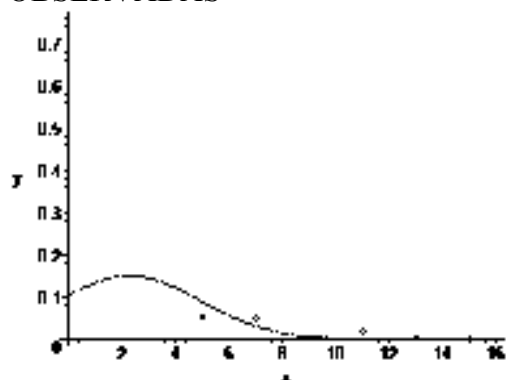
EXPONENCIAL

$$f(x) = \begin{cases} \frac{1}{3,4953} \cdot e^{\frac{-x}{3,4953}} & \text{para } x \geq 0, \beta > 0 \\ 0 & \text{e.o.c.} \end{cases} \quad \text{EQUAÇÃO 40}$$

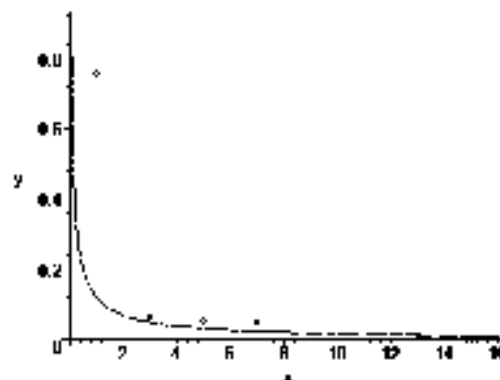
MODELO PROPOSTO

$$f(x) = \frac{x^2}{(0,2 + 0,1161 \cdot x)^{6,2550}} \quad \text{EQUAÇÃO 41}$$

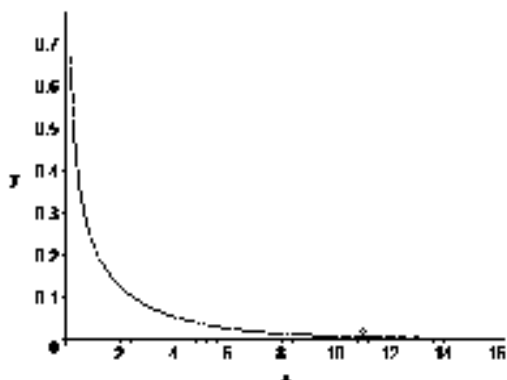
FIGURA 26 – REPRESENTAÇÃO GRÁFICA DAS FUNÇÕES CLÁSSICAS E A DESENVOLVIDA NESTE TRABALHO AJUSTADAS AOS DADOS DE REGENERAÇÃO NATURAL DE SASSAFRÁS E AS RESPECTIVAS FREQUÊNCIAS OBSERVADAS



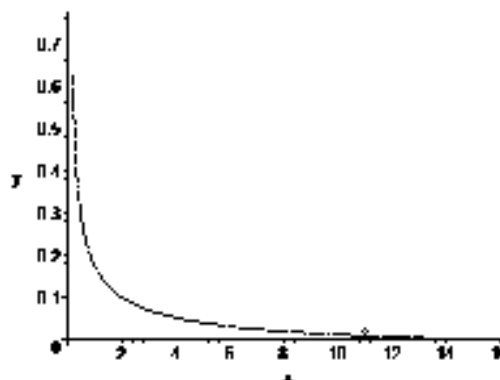
Normal



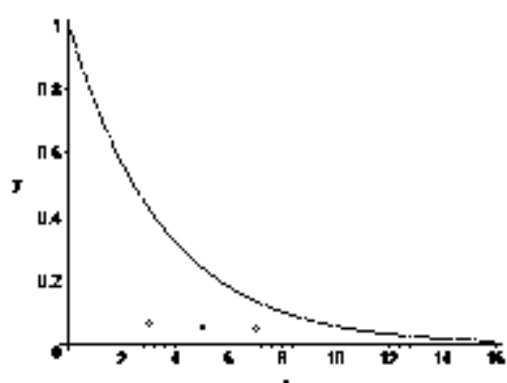
Weibull



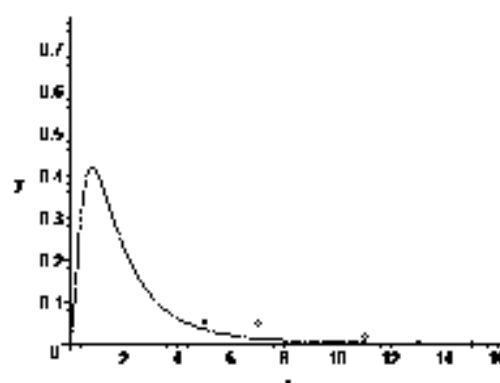
Gamma



Beta



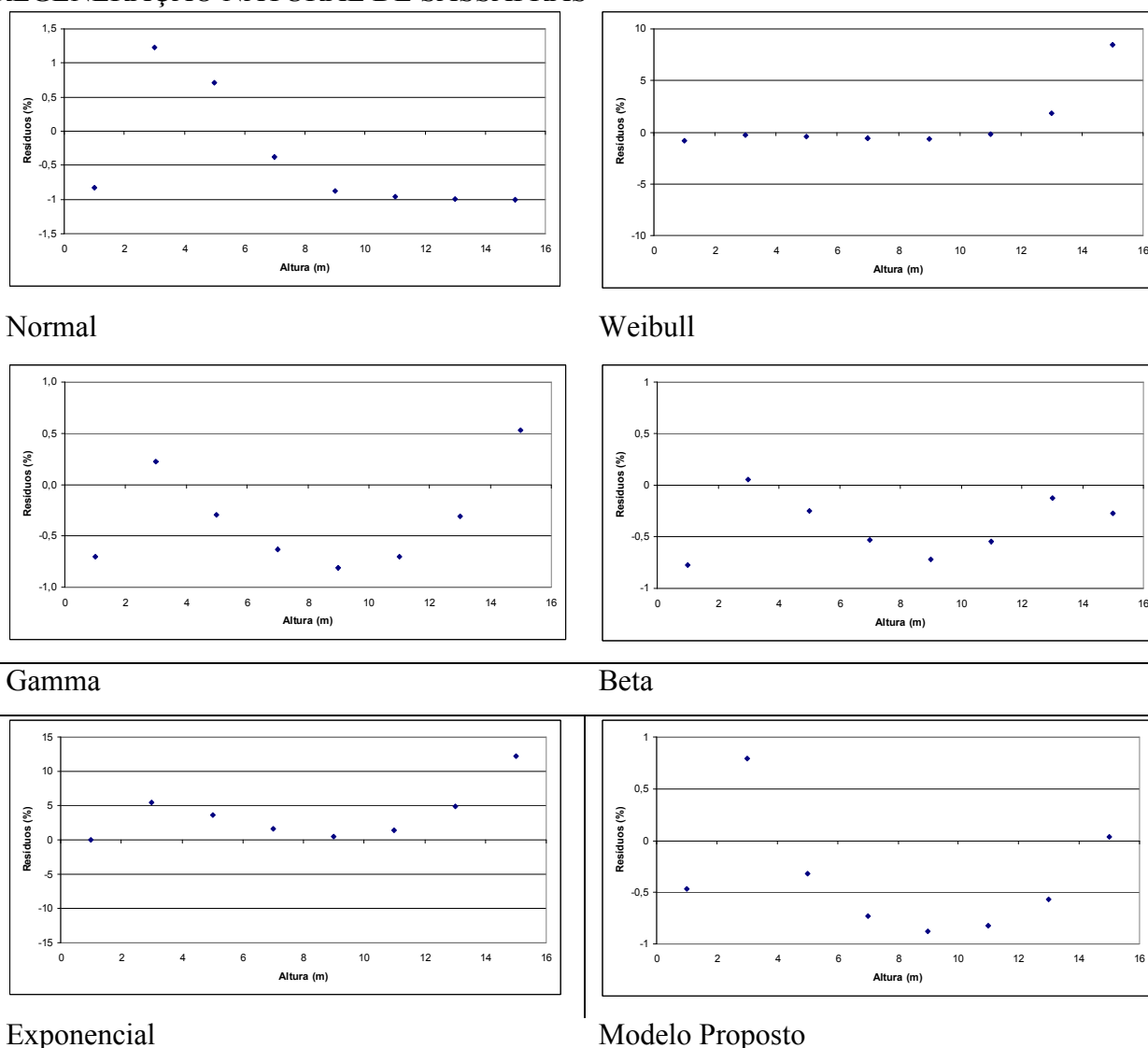
Exponencial



Modelo Proposto

Com o objetivo de verificar, visualmente, os ajustes dos modelos aos dados foram elaborados os gráficos de resíduos em percentagem (FIGURA 27).

FIGURA 27 – RESÍDUOS EM PORCENTAGEM DAS FUNÇÕES CLÁSSICAS E A DESENVOLVIDA NESTE TRABALHO AJUSTADAS AOS DADOS DE REGENERAÇÃO NATURAL DE SASSAFRÁS



Na Tabela 03 estão apresentados as médias, desvios padrões e coeficientes de variação calculados por meio das expressões desenvolvidas para cada modelo (Normal, Weibull, Gamma, Beta e Exponencial), inclusive o proposto neste trabalho. Na primeira linha estão apresentadas as estimativas dos parâmetros, obtidas a partir dos dados observados em campo.

Pode-se observar na Tabela 03 que os valores da média e do desvio padrão resultantes das distribuições Gamma, Beta e o proposto no presente trabalho resultaram em valores próximos aos observados. Como a Função de Gauss utiliza a estimativa da média e do desvio padrão como coeficientes, não tem sentido compará-la com as demais, quanto a esses dois parâmetros.

TABELA 03 – MÉDIA, DESVIO PADRÃO E COEFICIENTE DE VARIAÇÃO

Distribuição	Média (m)	Desvio padrão (m)	Coeficiente de variação
Observado	2,2913	2,6601	116,0958
Normal	2,2913	2,6601	116,0958
Weibull	5,1160	14,7395	288,1061
Gamma	2,2913	2,6601	116,0958
Beta	2,2912	2,6608	116,0958
Exponencial	3,4953	3,4953	1221,7134
Proposto	2,2909	2,7065	118,1419

A comparação entre os modelos foi feita por meio de R^2 , R^2 ajustados, erro padrão, teste de aderência de Kolmogorov-Smirnov, χ^2 e os desvios relativos, de acordo com a tabela 04.

TABELA 04 – AVALIAÇÃO DOS MODELOS

	Normal	Gamma	Beta	Weibull	Exponencial	Proposto
R^2	0,3509	0,9222	0,8976	0,7350	0,7759	0,9484
R^2 adj	0,2427	0,9093	0,8805	0,6290	0,7759	0,9096
Syx	0,2585	0,2186	0,2400	0,2869	0,1571	0,1805
Syx%	11,2825	9,5422	10,4741	12,5204	6,8554	7,8759
Kolmogorov	0,1248 **	0,1012 **	0,0966 **	0,1054 **	1,5309	0,1098 **
χ^2	4,7363	1,5247 **	2,1310 *	0,3306 **	0,5544 **	0,8275 **
Desvio relativo	0,5944	0,6199	0,6694	0,7270	-0,7063	0,0020
Desvio relativo 2	0,9188	0,8993	0,9374	0,9682	-0,4092	0,1054
Desvio relativo 3	0,9859	0,9723	0,9874	0,9958	-0,1834	0,2163
Desvio relativo 4	0,9975	0,9920	0,9973	0,9994	-0,0717	0,2940

Os valores tabelados para o teste de Kolmogorov-Smirnov para as probabilidades de 5% e 1%, respectivamente, são 0,576292 e 0,480833, enquanto que para o teste de χ^2 são 2,732637 e 1,646497, respectivamente.

Analisando a Tabela 04 pode-se afirmar que o modelo aqui desenvolvido resultou em um maior coeficiente de determinação múltiplo ajustado e menores desvios relativos de primeira e segunda ordem. Todos os modelos ajustaram-se aos dados, exceto o Modelo Exponencial, de acordo com o Teste de aderência de Kolmogorov-Smirnov, entretanto, segundo o Teste de χ^2 , somente a Curva de Gauss não foi considerado aderente.

7 - CONCLUSÕES

O presente estudo possibilitou afirmar que:

- a) Há uma grande variedade de modelos matemáticos que podem ser transformados em funções densidade de probabilidade;
- b) Existem diversas funções probabilísticas e seu processo de construção exige um conhecimento profundo dos conceitos das distribuições contínuas já existentes;
- c) A aplicação dos modelos clássicos aos dados de altura da regeneração natural do sassafrás não resultou em ajustes satisfatórios. A Distribuição Normal foi a única que apresentou moda, porém não foi aderente. Os demais modelos, por não apresentarem moda, não representaram a realidade dessa distribuição de frequências. A função Exponencial, considerada a mais própria para essas situações, resultou em um bom ajuste, entretanto, além de não apresentar moda, essa distribuição intercepta o eixo das ordenadas em um ponto β , o que significa que há β indivíduos com altura 0 e isso não é correto;

A análise do modelo proposto permite exarar as seguintes conclusões:

- 1. O ajuste do modelo atende ao propósito da pesquisa, que visou propor uma nova função densidade de probabilidade aplicável à regeneração natural;
- 2. O modelo é flexível, podendo assumir diversas formas, de acordo com a distribuição dos dados;
- 3. b desloca a função ao longo do eixo x ;
- 4. c altera a forma da função, variando sua dispersão ao longo do eixo x , e acumula toda a probabilidade em um único ponto, quando assume valor 0 (zero), se b for diferente de 0 (zero);
- 5. a e d também alteram a forma da função;
- 6. Todos os coeficientes contribuem para a média, a variância, a moda e para os pontos de inflexão, como pode ser observado nas expressões apresentadas acima;
- 7. O modelo proposto pode ser considerado uma função densidade de probabilidade, pois atende a todos os pré-requisitos necessários;
- 8. Ao comparar as principais distribuições utilizadas no meio florestal e a desenvolvida neste trabalho, verificou-se que essa se ajustou de forma satisfatória aos dados, podendo inclusive ser utilizada como função preditiva, como pode ser observado na discussão dos resultados;

9. Devido à flexibilidade do presente modelo, pode-se afirmar que esse pode ser aplicado a diversas distribuições de dados do setor florestal, bem como de outras áreas, cabendo aos pesquisadores sua aplicação;
10. O modelo proposto apresenta a vantagem de ser facilmente ajustado com o auxílio do *software Table Curve 2d*, em comparação com a maioria dos modelos clássicos;
11. Outra vantagem da função proposta nesse trabalho é a possibilidade de se calcular os parâmetros em função dos coeficientes;
12. Utilizando-se as expressões desenvolvidas, encontrou-se que:
 - a. A altura média estimada pelo modelo foi de 229 cm, com 271 cm de desvio padrão, ou seja, o coeficiente de variação nesse caso foi de 118,14%;
 - b. A moda da distribuição de alturas, 81 cm;
 - c. A função ajustada apresenta dois pontos de inflexão, 18,50 e 143,38 cm;

A comparação entre os modelos clássicos e o proposto neste trabalho permite afirmar que:

- I. Os Modelos Normal, Gamma, Beta e o aqui proposto foram os que mais se aproximaram da estimativa da média e desvio padrão;
- II. O modelo apresentado neste estudo resultou em um maior coeficiente de determinação múltiplo ajustado e menores desvios relativos de primeira e segunda ordem. Todos os modelos ajustaram-se aos dados, exceto o Modelo Exponencial, de acordo com o teste de aderência de Kolmogorov-Smirnov, entretanto, segundo o teste de χ^2 , somente a Curva de Gauss não foi considerado aderente; e
- III. A distribuição proposta neste trabalho resultou nos valores estimados mais próximos dos observados.

8 - REFERÊNCIAS

- ARCE, J. E.; GOMES, F. S.; SANQUETTA, C. R.; CRUZ, E. C. Utilização dos processos de difusão na avaliação e simulação precoces do crescimento de povoamentos de *Pinus taeda* L. **Cerne**. v. 4, n. 1, p. 154-170. 1998b.
- AYRES JR., F. **Cálculo diferencial e integral**: resumo da teoria, problemas resolvidos, problemas propostos. 2ª ed. São Paulo, SP: Poligráfica. P. 371, 1976. Coleção Schaum.
- BOYER, C. B. **História da matemática**. Ed. 2. Edgard Blücher. 1996.
- CALIJURI, M. L.; RÖHM, S. A. **Sistemas de Informações Geográficas**. Imp. Univ., Univ. Federal de Viçosa, Viçosa, 1995.
- BARROS, P. L. C. **Estudo das distribuições diamétricas da floresta do Planalto Tapajós - Pará**. 1980. 123f. Dissertação (Mestrado em Ciências Florestais) - Universidade Federal do Paraná, Curitiba, 1980.
- CARVALHO, J. O. P. **Análise estrutural da regeneração natural em floresta tropical densa na região do Tapajós no Estado do Pará**. Curitiba, 1982. 128 f. Dissertação (Mestrado em Engenharia Floresta) – Setor de Ciências Agrárias, Universidade Federal do Paraná.
- CARVALHO, P. E. R. **Espécies arbóreas brasileiras**. Brasília: Embrapa Informação Tecnológica; Colombo – PR: Embrapa Florestas, 2003. p.317
- CORVELLO, W. B. V. **Utilização de mudas da regeneração natural em reflorestamento com espécies nativas**. Curitiba, 1983, 105 f. Dissertação (Mestrado em Engenharia Florestal) – Setor de Ciências Agrárias, Universidade Federal do Paraná.
- COUTO, H. T. Z. **Apostila de inventário florestal**. Piracicaba: ESALQ, p.12, s.d.
- CRESTANA, C. S. M. Estrutura populacional de *Genipa americana* L. (RUBIACEAE) em mata riparia do rio Moji-Guaçu, em Conchal-SP. **Revista do Instituto Florestal**, São Paulo, v. 12, n.2, p. 105-117, 2000.
- EUES, H. **Introdução à história da matemática**. UNICAMP. 1995.
- FINOL, U. H. Nuevos parametros e considerarse en el analisis estructural de las selvas virgines tropicales. **Revista Forestal Venezolana**, Merida, v. 14, n. 21, p. 29-42, 1971.
- FINOL, U. H. Possibilidades de manejo silvicultural para las Reservas Forestales de la Region Occidental. **Revista Forestal Venezolana**, 12(17):81-107, 1969.
- FLEMMING, D. M.; GONÇALVES, M. B. **Cálculo A**: funções, limite, derivação, integração. 5ª ed, São Paulo, SP: Makron. P. 617, 1992.
- GMURMAN, V. E. **Teoria das probabilidades e estatística matemática**. Moscou: Mir, p. 139-150, 1983.
- GOMIDE, G. L. A. **Estrutura e dinâmica de crescimento de florestas tropicais**

primária e secundária no Estado do Paraná. Curitiba: Dissertação (Mestrado em Engenharia Florestal) - UFPR, 1997.

GUIDORIZZI, H. L. **Um curso de cálculo.** 2.ed. São Paulo: LTC, p. 579, 1987

GUIDORIZZI, H. L. **Um curso de cálculo.** 5.ed. São Paulo: LTC, p. 566, 1994.

HOSOKAWA, R.T. **Manejo e economia de florestas.** Roma: FAO, p. 125, 1986.

IFRAH, G. **História universal dos algarismos.** V. 1. Nova Fronteira. 1947.

INOUE, M. T. **Regeneração Natural – Seus Problemas e Perspectivas para as Florestas Brasileiras.** Curitiba: FUPEF-UFPR, Série Técnica n}1, 1979.

INOUE, M. T.; RODERJAN, C V., KUNIYOSHI, Y. S. **Projeto madeira do Paraná. Curitiba: Fundação de Pesquisas Florestais do Paraná,** 1984.

JOHNSON, N. L.; LEONE, F. C. **Statistics and experimental design in the engineering and the physical sciences.** V. 1, p. 523, 1964.

LEITHOLD, L. **Cálculo com geometria analítica.** Ed. 2. Harper & Row do Brasil. 1982.

LEITHOLD, L. **O cálculo com geometria analítica.** 3ª ed. São Paulo, SP: Harba, V.1, p. 390, 1994.

LONGHI, S. J. **A estrutura de uma floresta natural de *Araucaria angustifolia* (Bert.) O. Ktze, sul do Brasil.** 198 f. Dissertação (Mestrado). UFPR. Curitiba, 1980.

LORENZI, H. **Árvores brasileiras: manual de identificação e cultivo de plantas arbóreas nativas do Brasil.** v.1. São Paulo: Plantarum, 1998. p.127.

MALTAMO, M.; PUUMALAINEN, J.; PÄIVINEN, R. Comparison of Beta and Weibull Functions for Modelling Basal Area Diameter Distribution in Stands of *Pinus sylvestris* and *Picea abies*. **Scand. J. For. Res.**, v. 10, p. 284-295, 1995.

MARQUADT, D. W. An algorithm for least squares estimation of nonlinear parameters, **SIAM Journal**, V. 11, p. 431- 441, 1963

MAPLE 7.0, Waterloo Maple Inc.

MARIANO, G.; CRESTANA, C. S. M.; BATISTA, E. A.; GIONNOTTI, E.; COUTO, H. T. Z. Regeneração natural em área à margem de represa, no município de Piracicaba, Sp. **Revista do Instituto Florestal**, São Paulo, v. 10, n. 1, p. 81-93, jun. 1998.

MATTOS FILHO, A. Anatomia do lenho de *Ocotea pretiosa*. In: **Oleos Essenciais: tecnologia analítica, industrial e econômica** (15.: 1958, Rio de Janeiro). Boletim do Instituto de Óleos. Rio de Janeiro: Ministério da Agricultura, p.9

MEYER, P. L. **Probabilidade: aplicações à estatística.** Rio de Janeiro, p. 391, 1974.

MOOD, A. M.; GRAYBILL, F. A. **Introduccion a la teoria de la estadistica.** Madrid,

p. 536, 1978.

PENG, C. Growth and yield models for uneven-aged stands: past, present and future. **For. Ecol. Manage.**, v. 132, p. 259-279, 2000.

PISKOUNOV, N. **Cálculo diferencial e integral**. 16^a ed. Porto, Portugal: Lopes da Silva, V1, 1993.

PIZATTO, W. **Avaliação biométrica da estrutura e da dinâmica de uma Floresta Ombrófila Mista em São João do Triunfo - PR: 1999**. Curitiba: Dissertação (Mestrado em Engenharia Florestal) - UFPR, 1999.

REITZ, R.; KLEIN, R. M.; REIS, A. **Projeto madeira de Santa Catarina**: Herbário Barbosa Rodrigues/SUDESUL, Itajaí, 1978.

REITZ, R.; KLEIN, R. M.; REIS, A. **Projeto madeira do Rio Grande do Sul**. Itajaí: Herbário Barbosa Rodrigues/SUDESUL, 1988. p.411-416.

REZENDE, A. V. **Diversidade, estrutura, dinâmica e prognose do crescimento de um cerrado Sensu stricto submetido a diferentes distúrbios por desmatamento**. Curitiba: Tese (Doutorado em Engenharia Florestal) - UFPR, 2002.

ROLLET, B. La regeneración natural en bosque denso siempreverde de llanura de la Guayana Venezolana. Guayana Venezolana: **Centro de Documentación y Publicaciones del IFLAIC**, 1969. p. 39-73. (IFLAIC, 124).

ROSS, S. M. **A first course in probability**. 5^a ed., p. 514, 1998.

ROUSSAS, G. G. **A course in mathematical statistics**. 2^a ed. Londres: Adisson Wesley Publishing Company, p. 536, 1997.

SANQUETA, C. R.; NINOMIYA, O. K. Regeneration process em a Firhemlock Forest – Analysis of sampling population. In Trans. 103rd Jpn. **For. Soc.** 357-359, 1992.

SANQUETTA, C. R. **Fundamentos biométricos dos modelos de simulação florestal**. Curitiba: FUPEF Série Didática n. 8, 1996.

SEITZ, R. S. A regeneração natural na recuperação de áreas degradadas. In: Simpósio Sul-americano, 1., e Simpósio Nacional de Recuperação de Áreas Degradadas, 2., 1994, Foz do Iguaçu. **Anais...** Curitiba: FUPEF, p 103-110. 1994.

SILVA S. M.; BRITZ, R. M.; SOUZA, W. S. Análise da Regeneração Natural em Floresta Ombrófila Mista, São Mateus do Sul-PR. **Anais...** 1º Congresso Florestal Panamericano e 7º Congresso Florestal Brasileiro. Curitiba, 1993. Vol 1, 742.

SILVA, E. Q. **Nova função densidade de probabilidade aplicável à Ciência Florestal**. 2003. 98f. Tese (Doutorado em Engenharia Florestal) – Universidade Federal do Paraná, UFPR, Curitiba, PR, 2003

STEVENSON, W. J. **Estatística aplicada à administração**. P. 495, 1981.

STRUIK, D. J. **História concisa das matemáticas**. Ed. 2. Gradativa. 1987.

TABLE CURVE 2D v 3.03, Jandel, AISN Software Inc.

VANCLAY, J. K.; GILLICON, A. N.; KEENAN, R. J. Using functional attributes to quantify site productivity and growth patterns in mixed forests. **Forest Management**, v. 94, p. 149-163, 1997

VOLPATO, M. M. L. **Regeneração natural em uma floresta secundária no domínio de mata Atlântica**: uma análise fitossociológica. Viçosa: UFV, 1994. 123p. Dissertação (Mestrado em Engenharia Florestal).

ANEXO 1 – Linhas de comando para o desenvolvimento da Função densidade de probabilidade.

FUNÇÃO DENSIDADE DE PROBABILIDADE

Pacote de cálculo

> with(student);

[D, Diff, Doubleint, Int, Limit, Lineint, Product, Sum, Tripleint, changevar, completesquare, distance, equate, integrand, intercept, intparts, leftbox, leftsum, makeproc, middlebox, middlesum, midpoint, powsubs, rightbox, rightsum, showtangent, simpson, slope, summand, trapezoid]

Função

Função

> f(x):=(x^a)/((b+c*x)^d);

$$f(x) := \frac{x^a}{(b + c x)^d}$$

Propriedades da função

Valor da função quando $x = 0$

> subs(x=0,f(x));

0

Valor da função quando x tende a infinito

> limit(f(x),x=infinity);

$$\lim_{x \rightarrow \infty} \frac{x^a}{(b + c x)^d}$$

Como o valor de c é sempre maior que 2, por definição, o denominador tenderá a infinito mais rapidamente que o numerador, logo, quando x tende a infinito, a função tende a 0.

Função probabilística

Para que uma função seja probabilística, essa deve:

1. Ser integrável;
2. Tender a 0 (zero) quando x tender a infinito;
3. Ter sua integral igual a 1, quando x variar de 0 (zero) a infinito; e

4. Ter média e variância calculada utilizando-se os parâmetros.

Escrever integral indefinida da função

> Int(f(x),x);

$$\int \frac{x^a}{(b + c x)^d} dx$$

Resultado da integral indefinida da função

> int(f(x),x);

$$\frac{b^{(-d)} x^{(a+1)} \text{hypergeom}\left([a+1, d], [2+a], -\frac{x c}{b}\right)}{a+1}$$

Escrever integral da função, com x variando de 0 até infinito

> Int(f(x),x=0..infinity);

$$\int_0^{\infty} \frac{x^a}{(b + c x)^d} dx$$

Resultado da integral da função, com x variando de 0 até infinito

> int(f(x),x=0..infinity);

Definite integration: Can't determine if the integral is convergent.

Need to know the sign of --> a+1

Will now try indefinite integration and then take limits.

$$\frac{\left(\frac{c}{b}\right)^{(-a)} b^{(-d+1)} \Gamma(d-1-a) \Gamma(a+1)}{c \Gamma(d)}$$

Simplificação da integral da função, com x variando de 0 até infinito

> integral:=simplify(%);

$$integral := \frac{\left(\frac{c}{b}\right)^{(-a)} b^{(-d+1)} \Gamma(d-1-a) \Gamma(a+1)}{c \Gamma(d)}$$

Função probabilística, produto da função pelo inverso da integral

> F(X):=f(x)/integral;

$$F(X) := \frac{x^a c \Gamma(d)}{(b + c x)^d \left(\frac{c}{b}\right)^{(-a)} b^{(-d+1)} \Gamma(d-1-a) \Gamma(a+1)}$$

Escrever integral indefinida da função probabilística

> Int(F(X),x);

$$\int \frac{x^a c \Gamma(d)}{(b + c x)^d \left(\frac{c}{b}\right)^{(-a)} b^{(-d+1)} \Gamma(d-1-a) \Gamma(a+1)} dx$$

Resultado da integral indefinida da função probabilística

> int(F(X),x);

$$\frac{c x^{(a+1)} \left(\frac{c}{b}\right)^a \text{hypergeom}\left([a+1, d], [2+a], -\frac{x c}{b}\right) \Gamma(d)}{b (a+1) \Gamma(d-1-a) \Gamma(a+1)}$$

Escrever integral da função probabilística, com x variando de 0 até infinito

> Int(F(X),x=0..infinity);

$$\int_0^{\infty} \frac{x^a c \Gamma(d)}{(b + c x)^d \left(\frac{c}{b}\right)^{(-a)} b^{(-d+1)} \Gamma(d-1-a) \Gamma(a+1)} dx$$

Resultado da integral da função função probabilística, com x variando de 0 até infinito

> int(F(X),x=0..infinity);

ANEXO 2 – Linhas de comando para a obtenção das expressões que estimam a média, variância e desvio padrão.

Expectativa Matemática

Média

A média de uma função probabilística pode ser obtida por meio do cálculo dos momentos, ou seja, a média é o resultado da integral do produto da função probabilística por x, quando x varia de 0 (zero) a infinito.

Escrever a integral indefinida do produto da função probabilística por x:

> Int(F(X)*x,x);

$$\int \frac{x^a c \Gamma(d) x}{(b + c x)^d \left(\frac{c}{b}\right)^{(-a)} b^{(-d+1)} \Gamma(d-1-a) \Gamma(a+1)} dx$$

Resultado da integral indefinida do produto da função probabilística por x:

> int(F(X)*x,x);

$$\frac{c x^{(2+a)} \left(\frac{c}{b}\right)^a \text{hypergeom}\left([2+a, d], [3+a], -\frac{x c}{b}\right) \Gamma(d)}{b (2+a) \Gamma(d-1-a) \Gamma(a+1)}$$

Escrever a integral do produto da função probabilística por x, com x variando de 0 até infinito:

> Int(F(X)*x,x=0..infinity);

$$\int_0^{\infty} \frac{x^a c \Gamma(d) x}{(b + c x)^d \left(\frac{c}{b}\right)^{(-a)} b^{(-d+1)} \Gamma(d-1-a) \Gamma(a+1)} dx$$

Resultado da integral do produto da função probabilística por x, com x variando de 0 até infinito:

> int(F(X)*x,x=0..infinity);

$$\frac{b (a+1)}{c (-2+d-a)}$$

Média

Simplificação do resultado acima:

> media:=simplify(%);

$$media := -\frac{b(a+1)}{c(-d+2+a)}$$

Variância

A variância pode ser obtida pelo cálculo dos momentos, ou seja, a variância é a diferença entre o segundo momento e o quadrado do primeiro momento, a média. O segundo momento é igual a integral do produto da função probabilística pelo quadrado de x, quando x varia de 0 (zero) a infinito.

Escrever a integral indefinida do produto da função probabilística por x²:

> Int(F(X)*(x^2),x);

$$\int \frac{x^a c \Gamma(d) x^2}{(b + c x)^d \left(\frac{c}{b}\right)^{(-a)} b^{(-d+1)} \Gamma(d-1-a) \Gamma(a+1)} dx$$

Resultado da integral indefinida do produto da função probabilística por x²:

> int(F(X)*(x^2),x);

$$\frac{c x^{(3+a)} \left(\frac{c}{b}\right)^a \text{hypergeom}\left([3+a, d], [4+a], -\frac{x c}{b}\right) \Gamma(d)}{b(3+a) \Gamma(d-1-a) \Gamma(a+1)}$$

Escrever a integral do produto da função probabilística por x², com x variando de 0 até infinito:

> Int(F(X)*(x^2),x=0..infinity);

$$\int_0^{\infty} \frac{x^a c \Gamma(d) x^2}{(b + c x)^d \left(\frac{c}{b}\right)^{(-a)} b^{(-d+1)} \Gamma(d-1-a) \Gamma(a+1)} dx$$

Resultado da integral do produto da função probabilística por x^2 , com x variando de 0 até infinito:

> int(F(X)*(x^2),x=0..infinity);

$$\frac{b^2 (a+1) (2+a)}{c^2 (-3+d-a) (-2+d-a)}$$

Simplificação do resultado acima:

> X2:=simplify(%);

$$X^2 := \frac{b^2 (a+1) (2+a)}{c^2 (-d+3+a) (-d+2+a)}$$

Cálculo da variância (diferença entre o segundo momento e o quadrado do primeiro momento):

> X2-(media^2);

$$\frac{b^2 (a+1) (2+a)}{c^2 (-d+3+a) (-d+2+a)} - \frac{b^2 (a+1)^2}{c^2 (-d+2+a)^2}$$

Variância

Simplificação do resultado acima:

> variancia:=simplify(%);

$$variancia := - \frac{(a+1) b^2 (d-1)}{c^2 (-d+3+a) (-d+2+a)^2}$$

Desvio padrão

Raiz quadrada da variância:

> desvio_padrao:=sqrt(variancia);

$$desvio_padrao := \sqrt{- \frac{(a+1) b^2 (d-1)}{c^2 (-d+3+a) (-d+2+a)^2}}$$

ANEXO 3 – Linhas de comando para a obtenção das expressões que estimam a moda e os pontos de inflexão.

Função

> f(x):=(x^a)/((b+c*x)^d);

$$f(x) := \frac{x^a}{(b + c x)^d}$$

Moda

A moda representa o ponto, ou os pontos, de maior frequência, em uma fdp, e pode ser obtido a partir da primeira derivada da função em relação à x e igualando-se o resultado a 0 (zero).

Calcular a derivada primeira em relação a x da função:

> der:=diff(f(x),x);

$$der := \frac{x^a a}{x (b + c x)^d} - \frac{x^a d c}{(b + c x)^d (b + c x)}$$

Simplificando:

> simplify(%);

$$x^{(a-1)} a (b + c x)^{(-d)} - x^a (b + c x)^{(-d-1)} d c$$

Igualando-se a primeira derivada a 0 (zero):

> Moda:=solve(der=0,x);

$$Moda := -\frac{a b}{c (a - d)}$$

Pontos de inflexão

Os pontos de inflexão, se existirem, podem ser obtidos calculando-se o valor de x ao igualar a zero a segunda derivada de uma função.

Calcular a derivada segunda em relação a x da função:

> der2:=diff(f(x),x,x);

$$der2 := \frac{x^a a^2}{x^2 (b + c x)^d} - \frac{x^a a}{x^2 (b + c x)^d} - \frac{2 x^a a d c}{x (b + c x)^d (b + c x)} + \frac{x^a d^2 c^2}{(b + c x)^d (b + c x)^2} + \frac{x^a d c^2}{(b + c x)^d (b + c x)^2}$$

Igualando-se a segunda derivada a 0 (zero):

> der2_0:=solve(der2=0,x);

$$\begin{aligned}
 der2_0 &:= \frac{1}{2} \frac{(-2 a^2 + 2 d a + 2 a + 2 \sqrt{-d a^2 + d^2 a + d a}) b}{(-a + a^2 - 2 d a + d^2 + d) c}, \\
 &\frac{1}{2} \frac{(-2 a^2 + 2 d a + 2 a - 2 \sqrt{-d a^2 + d^2 a + d a}) b}{(-a + a^2 - 2 d a + d^2 + d) c}
 \end{aligned}$$

ANEXO 4 – Procedimentos para a obtenção dos coeficientes, utilizando-se o software

Table curve 2d.

1. Para importar os dados deve-se:
 - a. Clicar em “File”;
 - b. “Import...”;
 - c. Selecionar o arquivo com os dados;
2. Para ajustar a função:
 - a. Clica-se em “Process”;
 - b. “User functions...”;
 - c. Em “Function Name:” escrever um nome para a função, por exemplo “modelonovo”;
 - d. Em “Adjustable Parameter Count:” digitar o número de coeficientes a serem ajustados pelo modelo, neste caso “2”, pois os coeficientes “a” e “d” foram arbitrados como 5 e 16, respectivamente;
 - e. Em “Function:” escrever a função a ser ajustada, ou seja, $Y=(X^5)/((A0+A1*X)^{16})$, em que A0 e A1 são os coeficientes “b” e “c”, respectivamente;
 - f. A próxima etapa consiste em determinar os valores mínimos, de entrada e máximos de cada coeficiente a ser ajustado. Os valores mínimos e máximos para ambos os coeficientes são 0,00000001 e 10^{25} , respectivamente, que devem ser inseridos no programa como “0.00000001” e “1E+25”. Considera-se, ao determinar esses valores, que os coeficientes deverão estar contidos no intervalo de 0,00000001 até 10^{25} . O valor de entrada para “b” é 0,4286 (“0.4286”) e para “c” é 0,2071 (“0.2071”); e
 - g. Clica-se em “Fit UDFs” para ajustar a função e gerar o gráfico.